

UNIVERSITÉ DU QUÉBEC

THÈSE PRÉSENTÉE À
L'UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

COMME EXIGENCE PARTIELLE
DU DOCTORAT EN BIOPHYSIQUE ET BIOLOGIE CELLULAIRES

PAR
DAVID CHARBONNEAU

ÉLÉMENTS STRUCTURAUX ESSENTIELS À LA THERMOSTABILITÉ
DE NOUVELLES ENZYMES LIPOLYTIQUES

NOVEMBRE 2014

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire ou de cette thèse a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire ou de sa thèse.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire ou cette thèse. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire ou de cette thèse requiert son autorisation.

UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

Cette thèse a été dirigée par :

Marc Beauregard	Université du Québec à Trois-Rivières
Directeur de recherche, Ph. D.	Institution à laquelle se rattache l'évaluateur

Jury d'évaluation de la thèse :

Marc Beauregard, Ph. D.	Université du Québec à Trois-Rivières
Prénom et nom, grade	Institution à laquelle se rattache l'évaluateur

Simon Barnabé, Ph. D.	Université du Québec à Trois-Rivières
Prénom et nom, grade	Institution à laquelle se rattache l'évaluateur

Hugo Germain, Ph. D.	Université du Québec à Trois-Rivières
Prénom et nom, grade	Institution à laquelle se rattache l'évaluateur

Roberto A. Chica, Ph. D.	University of Ottawa
Prénom et nom, grade	Institution à laquelle se rattache l'évaluateur

Thèse soutenue le 11 décembre 2013

Chance is the only source of true novelty

Francis Crick

REMERCIEMENTS

Je me sens chanceux d'avoir pu travailler avec tous ces gens exceptionnels durant mes études de cycles supérieurs. J'aimerais premièrement remercier mon directeur de recherche, le professeur Marc Beauregard, qui a été un superviseur fantastique et un ami avec qui j'espère continuer de collaborer dans le futur. Je remercie aussi tout particulièrement ma collègue, la Dre Fatma Meddeb, avec qui j'ai partagé des moments de recherche, de discussion et de découverte mémorables. Je remercie aussi le Dr Yves Hurtubise avec qui ce projet a été initié. Merci aux membres du labo : Jessica Kelly Moisan, François Laframboise, Drs Yannick Hébert-Ouellet et Li Cui. Je garde d'excellents souvenirs de toutes ces années, que ce soit au labo, en congrès ou lors de nos fameuses réunions du vendredi après-midi. Je tiens à remercier particulièrement mon père Luc, ma mère Diane et ma sœur Janny-Claude qui m'ont supporté tout au long de mes études. Merci à mes amis, particulièrement à Jean-François Richard et Pascal Champagne avec qui j'ai partagé de longues discussions sur le sujet. Je tiens aussi à remercier le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) pour le soutien qui m'a été apporté durant mes études doctorales, les fonds de recherche du Québec – nature et technologies (FRQNT) ainsi que le regroupement PROTEO, qui m'ont permis de réaliser une collaboration de recherche à l'Université de Californie à Los Angeles (UCLA). À cet égard, je tiens à remercier particulièrement le professeur James Bowie pour m'avoir intégré à son équipe dans le cadre de cette collaboration ainsi que le Dr Tyler Paz Korman qui m'a accueilli dans le « Bowie lab » et initié à la cristallographie aux rayons X. Je garde d'excellents souvenirs de cette année passée au « Molecular Biology Institute » durant laquelle j'ai aussi eu la chance de côtoyer des chercheurs de renom tels que David Eisenberg et Richard E. Dickerson qui furent des modèles d'inspiration. Je remercie aussi le Dr Albert Berghuis de l'Université McGill, pour une collaboration de recherche en cristallographie. Finalement, je tiens aussi à remercier madame Catarina Leote Franco Pio pour son travail formidable ainsi que tout le personnel de l'Université du Québec à Trois-Rivières avec lequel j'ai collaboré durant mes études de cycles supérieurs.

AVANT-PROPOS

Le concept d'évolution des espèces proposé par Darwin dans son célèbre ouvrage « *On the origin of species* » a donné lieu à une révolution scientifique qui a pris son plein essor durant la deuxième moitié du vingtième siècle suite à la découverte de la structure moléculaire du matériel génétique et des mécanismes par lesquels ce matériel spécifie la synthèse et l'évolution des protéines soumises à la sélection naturelle. Ces avancées technologiques ont permis de fonder une nouvelle discipline, la biologie moléculaire, laquelle, couplée à la biologie structurale, est aujourd'hui au centre des activités de recherche en biologie. La face cachée du code génétique, c'est-à-dire la relation qui existe entre l'information unidimensionnelle contenue dans la séquence des gènes et la structure tridimensionnelle des protéines encodées par ces gènes, est un problème pour lequel une solution n'a pas encore vu le jour en ce début du 21^e siècle. Une meilleure compréhension de ce lien intime qui relie la séquence d'une protéine avec ses propriétés permettrait d'améliorer grandement l'efficacité de développement de nouveaux biocatalyseurs pour différentes applications biotechnologiques tant médicales qu'environnementales. Les différentes avancées sur l'étude de la structure des protéines et les différents problèmes reliés à la compréhension des mécanismes et des forces impliquées dans leur repliement et dans la stabilité de leur structure sont présentés dans l'introduction et discutés dans un contexte historique. Le problème du repliement est traité en parallèle à celui de la stabilité et différentes réflexions philosophiques y sont proposées.

Ce projet de thèse s'inscrit dans un contexte de chimie verte dans lequel nous tentons de développer des procédés industriels plus propres pour l'environnement par l'utilisation d'enzymes. La stabilité d'un biocatalyseur étant un critère d'application essentiel dans un procédé industriel, la compréhension des mécanismes par lesquels les enzymes acquièrent leur stabilité est donc nécessaire. Ce défi est ce qui m'a incité à entreprendre des recherches sur l'étude de la stabilité des enzymes dans le laboratoire du professeur Marc Beauregard. L'objectif initial et aussi le mandat qui m'a été donné à

l'hiver 2008 était d'identifier et de produire une nouvelle enzyme lipolytique thermostable afin d'étudier certains éléments structuraux responsables de sa thermostabilité en vue d'améliorer sa stabilité en milieu alcalin pour une application dans l'industrie des pâtes et papiers ainsi qu'en synthèse organique. Ce projet a été entamé dans le cadre du programme de maîtrise en biophysique et biologie cellulaires pour réaliser un passage direct au doctorat au printemps 2009.

Le projet de thèse présenté à l'hiver 2010 consistait à identifier par une approche rationnelle, certains éléments structuraux permettant de comprendre et de contrôler la stabilité d'une nouvelle enzyme lipolytique. L'étude de la stabilité des enzymes lipolytiques par évolution accélérée et par cristallographie aux rayons X m'a mené à aller entreprendre des recherches dans le laboratoire du professeur James Bowie à l'Université de Californie à Los Angeles durant l'année 2010. Ces travaux ont permis d'améliorer la thermostabilité et la résistance au méthanol d'une lipase par évolution accélérée pour la synthèse de biodiesel et de comprendre les éléments structuraux responsables de cette nouvelle stabilité par la résolution de la structure cristalline des variants obtenus au cours des rondes d'évolution accélérée. Cette thèse fait l'objet de cinq articles scientifiques publiés dans des journaux révisés par des pairs, les travaux de recherche ayant été réalisés entre 2008 et 2013 sous la direction du professeur Marc Beauregard à l'Université du Québec à Trois-Rivières. Les éléments structuraux identifiés comme responsables de la thermostabilité de deux enzymes lipolytiques distantes, mis en évidence par deux approches différentes, ont montré certaines similitudes quant au type d'interactions impliquées dans leur stabilité. Les différents résultats expérimentaux ont été présentés lors du séminaire doctoral à l'hiver 2013 et cette thèse a été défendue le 11 décembre 2013.

Articles scientifiques :

Article 1

Le premier article intitulé « Identification of thermophilic bacterial strains producing thermotolerant hydrolytic enzymes from manure compost » a fait l'objet d'une publication dans le journal *Indian Journal of Microbiology*, en 2011.

Article 2

Le deuxième article intitulé « A novel thermostable carboxylesterase from *G. thermodenitrificans* : Evidence for a new carboxylesterase family » a fait l'objet d'une publication dans le journal britannique *Journal of Biochemistry* (publié par la presse de l'Université d'Oxford) en 2010.

Article 3

Le troisième article intitulé « N-terminal purification tag alters thermal stability of the carboxylesterase EstGtA2 from *G. thermodenitrificans* by altering reversibility of thermal unfolding » a été publié dans le journal Américain *Protein and Peptide Letters* en 2012.

Article 4

Le quatrième article intitulé « Role of key salt bridges in thermostability of *G. thermodenitrificans* EstGtA2: Distinctive patterns within the new bacterial lipolytic enzyme family XV » est publié dans le journal *PLOS One (Public Library of Science)*.

Article 5

Le cinquième article intitulé « Dieselzymes: Développement of a stable and methanol tolerant lipase for biodiesel production by directed evolution » a fait l'objet d'une publication dans le journal *Biotechnology for Biofuels* en 2013 et est présenté en annexe.

RÉSUMÉ

L'objectif de cette thèse était d'identifier et de produire une nouvelle enzyme lipolytique thermostable et d'étudier certains éléments structuraux impliqués dans sa thermostabilité afin de mieux comprendre comment utiliser ces éléments pour améliorer sa stabilité en milieu alcalin en vue de son utilisation pour le contrôle des extractibles dans l'industrie des pâtes et papiers. Un compost organique a été utilisé comme source de microorganismes thermophiles pour la recherche de nouvelles enzymes lipolytiques thermostables. Quatre espèces bactériennes ont été isolées et identifiées : *Geobacillus thermodenitrificans*, *Bacillus smithii*, *Ureibacillus suwonensis* et *Aneurinibacillus thermoaerophilus*. Une nouvelle carboxylesterase de *G. thermodenitrificans* a été clonée chez *E. coli*, surexprimée, purifiée et caractérisée. Cette nouvelle enzyme lipolytique appelée EstGtA2 catalyse l'hydrolyse de la tributyrine et d'une large gamme d'esters aromatiques et aliphatiques entre 25-70 °C, de pH 6 à 10 avec une activité optimale à pH 8 et 50 °C, ce qui la rend intéressante pour différentes applications biotechnologiques. EstGtA2 montre un profil de stabilité thermique et de réversibilité conformationnelle dépendant du pH, sa renaturation *in vitro* est possible uniquement en milieu alcalin ce qui suggère que des interactions électrostatiques jouent un rôle prépondérant dans la stabilité de sa structure. Afin d'élucider certains éléments structuraux responsables de sa thermostabilité et de sa réponse au pH, la structure tridimensionnelle de EstGtA2 a été modélisée sur la base de la structure cristalline de MGL H-257, un proche homologue montrant 89 % d'identité de séquence. Ces enzymes appartiennent à un nouveau groupe d'enzymes lipolytiques (famille XV) et plus précisément à un clade phylogénétique particulier de la famille XV que nous avons appelé N'. Des alignements multiples de séquences et de structures ont permis d'identifier une série de ponts salins conservés au sein des enzymes appartenant au groupe N' et le rôle que jouent ces résidus de ponts salins dans la thermostabilité de EstGtA2 a été étudié par mutagenèse combinatoire. Les résultats montrent que certaines combinaisons de résidus basiques (Arg et Lys) impliqués dans la formation de ponts salins semi-enfouis sont essentielles à la thermostabilité de EstGtA2 alors qu'une combinaison particulière est essentielle à l'atteinte de sa conformation native. Nous avons identifié un pont salin particulier, conservé dans les familles XV et XIII, lequel montre une haute tolérance aux substitutions et aux renversements de polarité d'interactions (acide-base/base-acide). La mutation de ce pont a été utilisée pour contrôler la réponse au pH de EstGtA2 et améliorer sa stabilité en milieu alcalin. La coévolution de résidus chargés impliqués dans la formation de ponts salins au sein d'un groupe d'enzymes particulier constitue un élément de structure identifiable au niveau de la séquence qui peut être utilisée comme marqueur séquentiel d'une propriété comme la thermostabilité. Nous proposons l'utilisation de différents patrons de ponts salins conservés et distinctifs au sein des familles XV et XIII comme motifs de coévolution séquentielle pour l'annotation génomique et l'assignation de nouvelles séquences à ces familles d'enzyme lipolytiques.

Annexe

Lors d'un stage de spécialisation de 9 mois à UCLA, une approche d'évolution accélérée a été utilisée pour accroître la thermostabilité et la résistance au méthanol d'une lipase psychrophile de *Proteus mirabilis* (PML) pour la production de biodiesel. La plupart des mutations sélectionnées ayant émergé lors des rondes d'évolution accélérée sont impliquées dans la formation de nouveaux ponts hydrogènes et de réseaux de ponts salins, ce qui suggère que les interactions polaires et électrostatiques jouent un rôle particulièrement stabilisant dans l'environnement de faible constante diélectrique utilisé pour la synthèse de biodiesel (huile et alcool) dans lequel travaillent les enzymes lipolytiques. Le nouveau prototype appelé Dieselzyme 4 constitue la lipase la plus robuste à ce jour pour la synthèse de biodiesel et est présentement en attente de brevet.

Mots-clés : Carboxylesterase, lipase, thermostabilité, ponts salins, mutagenèse, composte, *Geobacillus thermodenitrificans*, *Aneurinibacillus thermoaerophilus*.

TABLE DES MATIÈRES

REMERCIEMENTS	iv
AVANT-PROPOS	v
RÉSUMÉ.....	viii
LISTE DES FIGURES	xvi
LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES	xix
LISTE DES SYMBOLES.....	xxi
CHAPITRE I	
INTRODUCTION.....	1
1.1 Les catalyseurs du vivant.....	1
1.2 Structure et évolution des enzymes	2
1.3 Niveaux de structures des protéines	4
1.4 Les enzymes lipolytiques.....	5
1.4.1 Les carboxylesterases et les lipases	6
1.4.2 Structure des enzymes lipolytiques.....	7
1.4.3 Site actif	9
1.4.4 Spécificité et cinétiques enzymatiques	10
1.4.5 Applications biotechnologiques des enzymes lipolytiques	12
1.4.6 Classification des enzymes lipolytiques	14
1.5 Les microorganismes thermophiles : une source d'enzymes thermostables	15
1.6 Repliement et stabilité des protéines	17
1.6.1 Détermination de la stabilité conformationnelle des protéines.....	20
1.6.2 Déterminants structuraux de la thermostabilité des protéines	24
1.6.3 Enzymes lipolytiques et thermostabilité	25
1.6.4 Les ponts salins : interactions clés chez les protéines thermophiles.....	26
1.6.5 Les interactions électrostatiques peuvent-elles jouer un rôle dans le repliement de certaines protéines?	28
1.7 Problème du repliement des protéines.....	29
1.7.1 Solution au problème du repliement : code ou énergie?.....	30

1.7.2	Modélisation de la structure des protéines et annotation génomique	32
1.7.3	Design de protéines et prédiction <i>ab initio</i>	34
1.7.4	Espace de séquence et explosion combinatoire	36
1.8	Évolution dirigée et étude de la thermostabilité par mutagenèses rationnelles ..	37
1.9	Objectifs.....	42
1.9.1	Objectif général.....	42
1.9.2	Objectifs spécifiques.....	42

CHAPITRE II

IDENTIFICATION OF THERMOPHILIC BACTERIAL STRAINS PRODUCING THERMOTOLERANT HYDROLYTIC ENZYMES FROM MANURE COMPOST		45
2.1	Contribution des auteurs	45
2.2	Résumé de l'article	45
2.3	Premier article scientifique	47
	Abstract.....	48
	Introduction	49
	Materials and methods.....	49
	Substrates and chemicals	49
	Source of microorganisms	50
	Isolation and culture of thermophiles	50
	PCR, RFLP and sequencing.....	50
	Analysis of 16S rRNA gene sequences	51
	Bacterial characterization	51
	Thermotolerant hydrolytic activity detection	51
	Results and discussion	52
	Identification of thermophilic bacteria	52
	Characterization of thermophiles.....	52
	Thermotolerant hydrolytic activities.....	53
	Acknowledgments	54
	References	55

CHAPITRE III

A NOVEL THERMOSTABLE CARBOXYLESTERASE FROM <i>GEOBACILLUS THERMODENITRIFICANS</i>: EVIDENCE FOR A NEW CARBOXYLESTERASE FAMILY.....	63
3.1 Contribution des auteurs	63
3.2 Résumé de l'article	63
3.3 Deuxième article scientifique	65
Abstract.....	66
Introduction	67
Materials and methods.....	67
Isolation of bacteria from environmental samples.....	67
Strains, plasmids and culture conditions.....	68
Preparation of genomic DNA	68
Polymerase chain reaction	68
Cloning of PCR products.....	69
DNA sequencing and bioinformatics analysis.....	69
Structural modelling	70
Protein expression.....	70
Protein purification and determination	70
Enzyme assays	71
Chain length specificity, optimal pH and temperature	71
Thermal stability	72
Results	72
Sequence of EstGtA2.....	72
Expression and purification of EstGtA2.....	73
Characterization of EstGtA2.....	73
Sequence comparison of EstGtA2	74
Structure modelling of EstGtA2	75
Phylogenetic tree of carboxylesterases	76
Discussion.....	76
Conclusion	79
Acknowledgments	80

Funding	80
References	81

CHAPITRE IV

N-TERMINAL PURIFICATION TAG ALTERS THERMAL STABILITY OF THE CARBOXYLESTERASE ESTGTA2 FROM *G. THERMODENITRIFICANS* BY IMPAIRING REVERSIBILITY OF THERMAL UNFOLDING.....

	94
4.1 Contribution des auteurs	94
4.2 Résumé de l'article	94
4.3 Troisième article scientifique	96
Abstract	97
Introduction	97
Materials and methods	99
Buffers and chemicals	99
DNA cloning and protein expression	99
Protein purification and cleavage of the N-terminal tag	99
Enzyme activity assay	100
CD measurements	100
Conformational and thermodynamics (unfolding/refolding)	101
Results	102
Expression and purification of the recombinant EstGtA2	102
Thermal stability of EstGtA2-tagged and -cleaved	102
Conformational and thermodynamic analysis (unfolding/refolding)	103
Conclusion	104
Acknowledgments	105
References	106

CHAPITRE V

ROLE OF KEY SALT BRIDGES IN THERMOSTABILITY OF *G. THERMODENITRIFICANS* ESTGTA2: DISTINCTIVE PATTERNS WITHIN THE NEW BACTERIAL LIPOLYTIC ENZYME FAMILY XV

	113
5.1 Contribution des auteurs	113
5.2 Résumé de l'article	113
5.3 Quatrième article scientifique	115

Summary.....	116
Introduction	117
Materials and Methods	120
Strains and plasmids	120
Site-directed mutagenesis	120
Protein expression and purification	121
Enzyme activity measurements	121
CD spectroscopy	122
Fluorescence spectroscopy	123
<i>In silico</i> analyses.....	123
Results	124
pH-dependent stability of EstGtA2.....	124
Structural model of EstGtA2	125
Multiple conserved salt bridges in EstGtA2	125
Salt bridges as hallmarks of the N' subfamily	127
Combinatorial alanine-shaving of salt bridge-forming residues.....	128
Misfolding of M4c	130
Activation and specificity modification by R37A	131
Modification of the (<i>i</i> - 2, <i>i</i> - 4) interloop salt bridge.....	132
Discussion.....	133
A 15 th family of bacterial lipolytic enzymes.....	133
Role of key salt bridges in thermostability of EstGtA2.....	135
Conclusion	138
Supporting information.....	138
Acknowledgements	140
References	141
Supporting information.....	160
CHAPITRE VI	
RÉSULTATS ET DISCUSSION	171
6.1 Approche choisie : mutagenèses rationnelles	172
6.2 EstGtA2 : une nouvelle enzyme lipolytique thermostable	172

6.3	Structure de EstGtA2	175
6.3.1	Une 15 ^e famille d'enzymes lipolytiques bactériennes (famille XV)	177
6.3.2	Le groupe N', un sous-clade de la nouvelle famille XV	178
6.3.3	Rôle des ponts salins conservés dans la thermostabilité de EstGtA2	179
6.3.4	Rôle des ponts salins conservés dans le repliement de EstGtA2	183
6.3.5	Modification d'un pont salin clé en position (<i>i</i> - 2, <i>i</i> - 4).....	186
6.3.6	Modifications d'autres ponts salins dans EstGtA2	187
6.3.7	Perspectives	188
CHAPITRE VII		
CONCLUSION.....		190
RÉFÉRENCES BIBLIOGRAPHIQUES.....		191
ANNEXE A		
DIESELZYMES: DEVELOPMENT OF A STABLE AND METHANOL TOLERANT LIPASE FOR BIODIESEL PRODUCTION BY DIRECTED EVOLUTION		202
A.1	Contribution des auteurs	202
A.2	Résumé de l'article	203
A.3	Cinquième article scientifique	204
A.4	Conclusion générale.....	218
ANNEXE B		
FIGURES SUPPLÉMENTAIRES		219

LISTE DES FIGURES

Figure		Page
1.1	Synthèse des protéines dans la cellule	3
1.2	Structures secondaires des protéines.....	3
1.3	Structure tertiaire de la myoglobine.....	4
1.4	Réactions catalysées par les carboxylesterases et les lipases.....	6
1.5	Topologie structurelle du repliement α/β hydrolase.....	7
1.6	Repliement tridimensionnel des α/β hydrolases.....	8
1.7	Mécanisme d'hydrolyse d'une liaison ester.....	9
1.8	Différence de cinétique enzymatique entre les carboxylesterases et les lipases.....	10
1.9	Topologie structurelle des lipases et éléments additionnels	11
1.10	Structure du couvercle « lid »	11
1.11	Structure des carboxylesterases	12
1.12	Principaux esters constituant les extractibles des tiges et du bois	13
1.13	Réaction de synthèse de biodiesel catalysée par les lipases	14
1.14	Phylogénie du vivant.....	16
1.15	Principaux déterminants structuraux de la stabilité des protéines	20
1.16	Spectres de dichroïsme circulaire en UV lointain.....	21
1.17	Relaxation par le solvant.....	22
1.18	Formation d'un pont salin.....	27
1.19	Repliement cotraductionnel des protéines	29
1.20	Repliement et dénaturation réversible des protéines	31
1.21	L'unité peptidique et les degrés de liberté permis	31
1.22	Schématisation théorique des différents paysages énergétiques associés aux différents états des protéines	32
1.23	Évolution par approche rationnelle et par évolution accélérée	41

1.24	Schéma hypothétique pour l'identification de résidus contrôlant la thermostabilité d'une nouvelle enzyme lipolytique thermostable à potentiel d'application industrielle	43
1.25	Schéma hypothétique pour l'identification de résidus contrôlant la thermostabilité de deux nouvelles enzymes lipolytiques thermostables à potentiel d'application industrielle	44
6.1	Réversibilité conformationnelle de EstGtA2	174
6.2	Cristal de EstGtA2	176
6.3	Photographie du patron de diffraction des rayons X de EstGtA2.....	176
6.4	Structure tridimensionnelle de EstGtA2	177
6.5	Structure des ponts salins étudiés dans EstGtA2	179
6.6	Localisation des substitutions alanine multiples.....	181
6.7	Impact des substitutions sur la stabilité EstGtA2	182
B.1	Table des codons (code génétique)	219
B.2	Liste des principales enzymes utilisées en industrie.....	220
B.3	Structure des acides aminés.	221
B.4	Vecteur utilisé pour le clonage et l'expression de EstGtA2	222
B.5	Double digestion du plasmide recombinant pQE31/estGtA2	223
B.6	Amplification de l'ORF de EstGtA2	224
B.7	Vecteur utilisé pour le sous-clonage et l'expression de EstGtA2	225
B.8	Colonies PCR.....	226
B.9	Séquence de EstGtA2 recombinante.....	227
B.10	Mesure des paramètres thermodynamiques de dénaturation	228
B.11	Dénaturation thermique de EstGtA2 à pH 8 (dichroïsme circulaire)	229
B.12	Dénaturation thermique de EstGtA2 à pH 8 (fluorescence)	230
B.13	Dénaturation thermique de EstGtA2 à différents pH.....	231
B.14	Réversibilité conformationnelle partielle de EstGtA2 à pH 8	232
B.15	Réversibilité conformationnelle dépendante du pH.....	233
B.16	Ponts salins étudiés dans EstGtA2.....	234
B.17	Séquences des mutants (R/K → A) combinatoires.....	235
B.18	Impact des substitutions alanine sur la stabilité conformationnelle de EstGtA2.....	236
B.19	Séquences des mutants H→R et X→C.....	237

B.20	Gel d'électrophorèse de protéines SDS-PAGE.....	238
B.21	Courbe de dénaturation thermique de PML et de M5.....	239
B.22	Cristallisation de M5.....	240
B.23	Patron de diffraction du cristal orthorhombique de M5.....	240
B.24	Changements structuraux dans le mutant M5.....	241
B.25	Activité de transestérification de PML et des différents mutants (première ronde d'évolution accélérée) mesurée par TLC	242
B.26	Spectre électromagnétique de la lumière	243
B.27	Diagramme de Jablonski.....	244

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

ADN	Acide désoxyribonucléique
Ala/ A	Alanine
Arg/ R	Arginine
ARN	Acide ribonucléique
ARNr	ARN ribosomal
Asp/ D	Acide aspartique
Asn/ N	Asparagine
BLAST	Basic local alignment search tool
CD	Circular dichroism
DMC	Double mutant cycle
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
FID	Flame ionization detector
GC	Gas chromatography
Glu/ E	Acide glutamique
Gln/ Q	Glutamine
His/ H	Histidine
IPTG	Isopropyl- β -D-1-thiogalactopyranoside
IMAC	Immobilized metal affinity chromatography
LB	Luria Bertani
Lys/ K	Lysine
MEGA	Molecular evolutionary genetics analysis
MSA	Multiple sequences alignment
MWCO	Molecular weight cut-off
Ni-NTA	Nickel-nitriloacetic acid

N-J	Neighbour-joining
OD	Optical density
ORF	Open reading frame
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PML	<i>Proteus mirabilis</i> lipase
<i>p</i> NP	Para-nitrophenol
RFLP	Restriction fragments length polymorphism
RMS	Root-mean square
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Ser/ S	Serine
TLC	Thin layer chromatography
Trp/ W	Tryptophane
X-gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside

LISTE DES SYMBOLES

Å	Ångström
cal	Calorie
°C	Degré centigrade
ΔS	Différentiel d'entropie
ΔH	Différentiel d'enthalpie
ΔG	Différentiel d'énergie libre
g	Gramme
J	Joule
K	Kelvin
kcal	Kilocalorie
k_{cat}	Constante catalytique
kDa	Kilodalton
K_M	Constante de Michaelis-Menten
L	Litre
λ	Longueur d'onde
M	Molaire
mg	Milligramme
mL	Millilitre
mM	Millimolaire
nm	Nanomètre
[\odot]	Éllipticité
T_m	Température de fusion
U	Unité d'activité enzymatique
μg	Microgramme
μL	Microlitre
μM	Micromolaire

CHAPITRE I

INTRODUCTION

1.1 Les catalyseurs du vivant

Les enzymes sont des macromolécules sélectives qui se caractérisent par un énorme pouvoir catalytique, pouvant accélérer de façon spécifique les réactions chimiques de la cellule à des taux de plus de 10^{16} fois ceux des niveaux non catalysés sans être elles-mêmes modifiées par la réaction. Elles sont les protéines responsables des transformations biochimiques au sein des cellules des organismes vivants et au centre de l'organisation du métabolisme et de la régulation des processus physiologiques. La spécificité des enzymes à catalyser des réactions chimiques précises est à la base de cette organisation et fait de ces molécules biologiques des catalyseurs de choix pour différentes applications biotechnologiques modernes. Les enzymes lipolytiques se spécialisent quant à elles dans le clivage ou la formation de liaisons esters et occupent une importante part du marché vu leurs nombreuses applications industrielles. La stabilité d'un biocatalyseur étant un important critère d'application en industrie, la recherche de nouvelles enzymes thermostables ou la modification d'enzymes connues en vue d'améliorer leur stabilité nécessite une meilleure compréhension des mécanismes par lesquels ces enzymes acquièrent leur stabilité. Le rythme auquel de nouveaux génomes de microorganismes sont séquencés et déposés dans les bases de données dépasse largement celui par lequel la structure des protéines encodées par ces nouveaux génomes peut être élucidée. Une meilleure compréhension de la relation qui existe entre la séquence d'acides aminés et la structure tridimensionnelle des protéines et l'identification d'éléments structuraux déterminants de l'évolution d'une propriété comme la thermostabilité permettrait de mieux cibler la recherche de nouveaux biocatalyseurs et d'adapter leurs propriétés pour l'industrie.

1.2 Structure et évolution des enzymes

L'histoire moderne des enzymes commence en 1831 quand l'auteur allemand Erhard Friedrich Leuchs remarque qu'un principe actif contenu dans la salive est capable d'hydrolyser l'amidon. En 1833, Anselme Payen et Jean-François Persoz décrivent un principe similaire isolé de l'orge en germination, qu'ils nomment diastase, dérivé du Grec *diastasis*, qui signifie séparation. Eduard Buchner démontre en 1897 que les microorganismes qui provoquent la fermentation alcoolique ne le font qu'indirectement puisque celle-ci peut survenir à partir d'extraits de levure, dépourvus de cellules vivantes. Le mot enzyme signifie littéralement « *in yeast* » (dans la levure). Émile Duclaux propose ensuite d'utiliser le suffixe -ase pour la nomenclature des enzymes. En 1902, Emil Fischer et Franz Hofmeister conclurent indépendamment que les protéines sont formées de longues chaînes d'acides aminés et que celles-ci peuvent être isolées et cristallisées et en 1926, James Sumner démontre la nature protéique des enzymes par ses travaux sur l'uréase. Le développement des techniques de cristallographie et d'analyse par rayons X ayant d'ailleurs permis d'élucider la structure de la molécule d'ADN en 1953 rendit possible d'étudier la structure complexe des protéines et d'expliquer le mécanisme catalytique des enzymes sur la base de leur structure tridimensionnelle (Bernal et Crowfoot 1933, Bragg et Perutz 1954, Crick 1956). C'est alors qu'en 1960, John Kendrew élucide pour la première fois la structure moléculaire d'une protéine, la myoglobine. Les travaux subséquents des membres du « RNA Tie Club¹ » et ceux de Marshall Nirenberg permirent d'élucider la nature du code génétique², de démontrer comment les protéines sont encodées et synthétisées dans la cellule et de suggérer un mécanisme pour leur évolution (Watson et Crick, 1953; Crick 1958, 1961, 1988; Brenner 1961; Nirenberg *et al.*, 1965). Le principe de sélection naturelle avancé par Darwin pouvait enfin être expliqué sur une base moléculaire.

¹ Le RNA Tie Club est un regroupement de 20 scientifiques de différentes disciplines qui avaient comme même objectif de comprendre comment l'ADN spécifie la synthèse des protéines. Fondé en 1954 par James Watson et Georges Gamow, il regroupait entre autres Francis Crick, Sydney Brenner, Richard Feynman, Erwin Chargaff, Leslie Orgel et Max Delbrück.

² Le code génétique réfère à la correspondance entre les séquences d'acides nucléiques (ADN et ARN) et les séquences d'acides aminés (protéines). La séquence d'acides aminés de chaque protéine synthétisée par une cellule est spécifiée par la séquence d'ADN du gène correspondant dans son génome. Une mutation dans la séquence codante d'un gène entraîne une mutation dans la protéine encodée par ce gène. Le tableau du code génétique est présenté en annexe B.

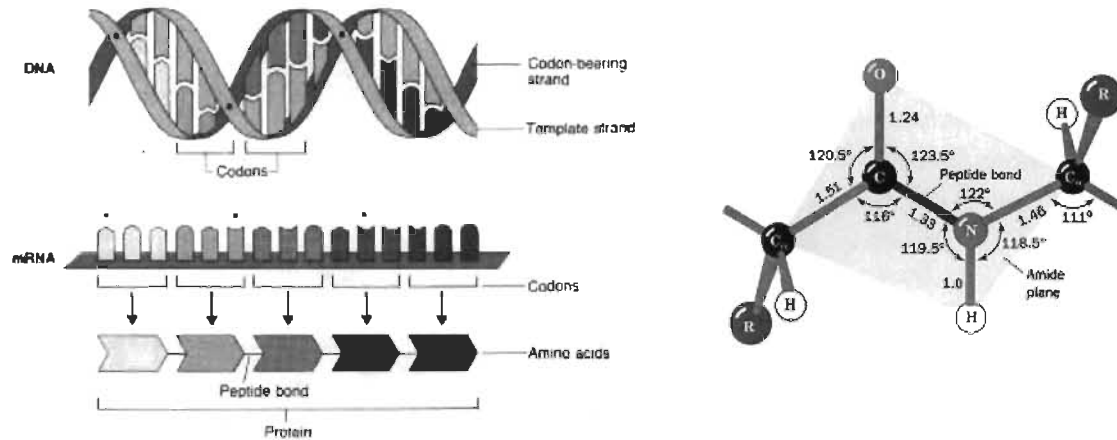


Figure 1.1 Synthèse des protéines dans la cellule.

Une séquence d'ADN (gène) est transcrite en ARNm par une polymérase et traduite en séquence d'acides aminés (protéine) par le ribosome. À gauche, une représentation de la correspondance entre la séquence nucléotidique d'un gène et la séquence d'acides aminés de la protéine correspondante. À droite, représentation du lien peptidique (planaire), la formation d'un lien amide covalent entre le groupe carboxyl et amine de deux acides aminés. Les angles dièdres ϕ et ψ sont formés entre N-C α et C-C α respectivement.

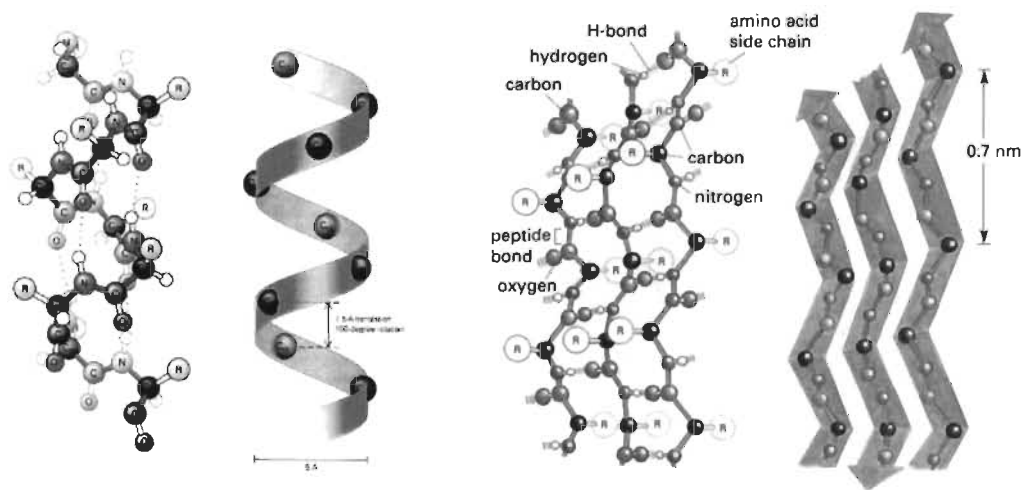


Figure 1.2 Structures secondaires des protéines.

La chaîne polypeptidique d'une protéine peut se stabiliser en adoptant deux types de structures. À gauche, la structure en hélice alpha est stabilisée par la formation de ponts hydrogènes entre les groupes CO et NH de la chaîne polypeptidique des résidus n et $n + 4$. À droite, structure en feuillet bêta stabilisée par des interactions hydrogènes entre les groupes CO et NH d'une chaîne polypeptidique se repliant sur elle-même.

1.3 Niveaux de structures des protéines

Avant même l'élucidation de la première structure tridimensionnelle d'une protéine, Linderstøm-Lang proposa de diviser la structure des protéines en quatre niveaux hiérarchiques (Linderstrøm-Lang 1951). La séquence d'acides aminés d'une protéine, c'est-à-dire l'ordre dans lequel ceux-ci sont agencés forme la structure primaire alors que les arrangements locaux des acides aminés de la séquence forment quant à eux les éléments de structures dites secondaires. L'hélice alpha et le feuillet bêta constituent les deux principaux types de structures secondaires et avaient été prédits par Linus Pauling environ dix ans avant l'élucidation de la structure cristallographique de la myoglobine (Pauling 1948, 1951, 1960). Le repliement tridimensionnel de ces éléments de structures secondaires forme la structure dite tertiaire qui confère la fonction biologique à une protéine et l'oligomérisation ou l'association de différentes unités tertiaires forme la structure quaternaire de certaines protéines (Figures 1.2 et 1.3).

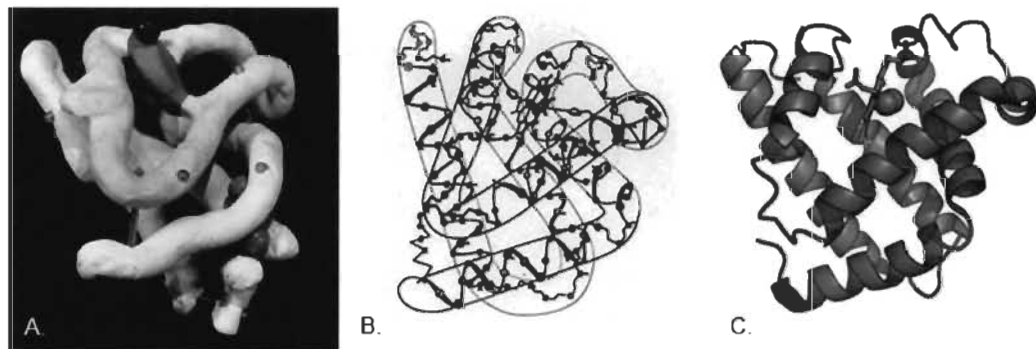


Figure 1.3 Structure tertiaire de la myoglobine.

A et B. Modèle de la myoglobine du cachalot réalisé par John Kendrew en 1958 basé sur les cartes de densité électronique issues du patron cristallographique à une résolution de 6 Å. (Kendrew *et al.*, 1958, 1960).
C. Structure de la myoglobine à une résolution de 2 Å (PDB no. 1MBN).

Le premier constat observé lors de l'élucidation de la structure de la myoglobine est l'irrégularité déconcertante de sa structure tertiaire. La chaîne polypeptidique adopte une structure globulaire irrégulière, mais nécessairement unique. L'absence de symétrie dans la structure tertiaire d'une protéine soulève différentes questions sur le mode de repliement unique de sa chaîne polypeptidique. Une des évidences principales de cette

unicité structurale est que les protéines peuvent former des cristaux, hautement symétriques, ceux-ci nécessitent l'arrangement périodique d'un objet de structure définie, ce qui permet de résoudre leur structure par diffraction des rayons X. De plus, la spécificité des enzymes et des autres protéines de la cellule sont le résultat de l'unicité de leur structure. Un autre constat a été fait en analysant la structure de la myoglobine, l'absence de liaisons chimiques entre les résidus de l'intérieur de la structure globulaire. Quelles sont les forces qui maintiennent la structure globulaire et surtout celles qui provoquent le repliement unique de la chaîne polypeptidique? En fait, les résidus d'acides aminés polaires d'une chaîne polypeptidique font contact avec l'eau du solvant alors que les résidus non polaires se regroupent formant un cœur hydrophobe à l'intérieur de la protéine. L'effet hydrophobe semble être la force responsable du repliement de la chaîne polypeptidique d'une protéine, mais comment la structure unique d'une protéine est-elle déterminée et quel est le rôle des autres types d'interactions dans ce processus qui confère la fonction et la stabilité d'une protéine?

1.4 Les enzymes lipolytiques

Les enzymes se divisent en six classes distinctes sur la base du type de réaction catalysée selon l'« *Enzyme Commission*³ ». Les hydrolases forment la troisième classe (EC 3). Ces enzymes catalysent l'hydrolyse ou la synthèse de liaisons C-O, C-N ou C-C dans différents types de substrats. Les enzymes lipolytiques sont des hydrolases qui catalysent l'hydrolyse ou la synthèse de liaisons esters (C-O) dans les huiles et autres esters aliphatiques, elles sont impliquées dans le métabolisme des lipides. Les enzymes lipolytiques sont ubiquitaires dans la nature et on les retrouve dans les trois règnes du vivant jouant des rôles physiologiques très diversifiés. Chez les mammifères par exemple, elles sont notamment impliquées dans la digestion, l'absorption et l'utilisation des lipides et des vitamines, dans la transformation métabolique de médicaments naturels et artificiels ainsi qu'au niveau de la détoxification hépatique et de l'utilisation de neurotransmetteurs (Hui et Howles, 2002) pour ne citer que quelques exemples. Chez

³ « Enzyme Commission number » est une classification numérique des enzymes basées sur le type de réaction chimique catalysée.

les microorganismes, elles jouent un rôle dans l'utilisation de sources de carbone et dans la détoxification. Le rôle métabolique que jouent ces enzymes chez les microorganismes reste encore obscur (Liu *et al.*, 2004).

1.4.1 Les carboxylesterases et les lipases

Les carboxylesterases (EC 3.1.1.1) et les lipases (EC 3.1.1.3) sont des enzymes solubles qui catalysent le clivage ou la formation de liaisons esters au niveau de différents types de molécules, notamment différents types d'esters aliphatiques, aromatiques ou de triglycérides. Elles peuvent aussi catalyser l'alcoololyse de liens esters et différentes réactions de transesterification entre des huiles et des alcools. Ces enzymes sont très versatiles et peuvent catalyser différentes réactions sur une large gamme de substrats. Bien que la spécificité de ces enzymes soit plus souple, la longueur de chaîne acylée est une des principales différences de spécificité entre les carboxylesterases qui préfère les substrats partiellement hydrosolubles à courtes chaînes (> 10 carbones) et les lipases qui nécessitent des substrats liposolubles à longues chaînes (< 10 carbones) et sont activées à l'interface lipide-eau.

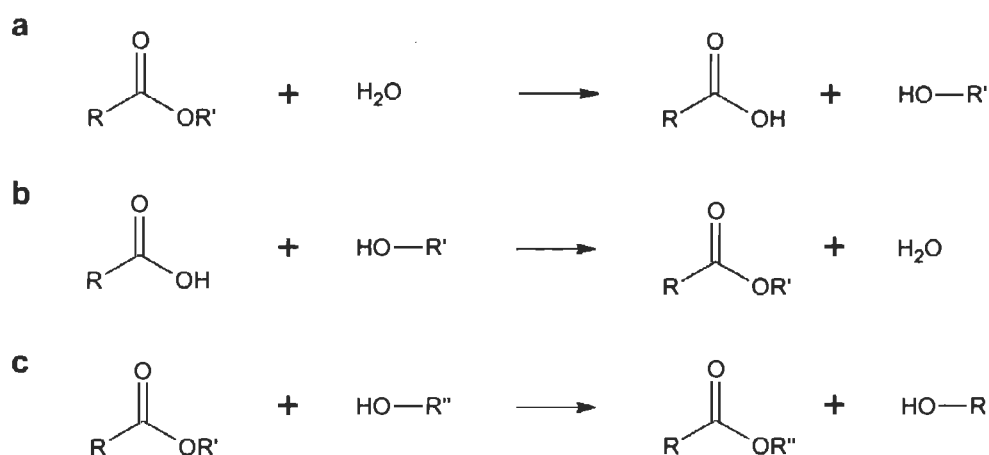


Figure 1.4 Réactions catalysées par les carboxylesterases et les lipases.
Réaction d'hydrolyse (a), de synthèse (b) et de transesterification (c).

1.4.2 Structure des enzymes lipolytiques

Les enzymes lipolytiques adoptent le repliement alpha/bêta hydrolase qui consiste en un domaine principal formé par un feuillet bêta central parallèle ou mixte, entouré par des hélices alpha (Ollis *et al.*, 1992). La topologie du repliement alpha/bêta hydrolase « canonique » est représentée à la figure 1.5. Ce repliement consiste en un feuillet bêta central composé de huit brins presque tous parallèles, seul le deuxième brin est antiparallèle, ce feuillet étant entouré d'hélices, habituellement six. Le feuillet central présente une torsion main-gauche, avec le premier et le dernier brin bêta se croisant à un angle de près de 90°. Certaines de ces hélices peuvent être absentes, seule l'hélice αC semble être conservée dans toutes les protéines adoptant ce repliement, jouant un rôle clé dans le positionnement de la sérine nucléophile située dans la boucle entre le brin $\beta 5$ et l'hélice αC (Figure 1.6). Cette boucle conservée est appelée le coude nucléophile et est normalement formée d'un pentapeptide conservé, G-X-S-X-G, au niveau duquel la sérine est présentée. Les deux autres résidus de la triade catalytique Asp/Glu et His sont aussi retrouvés au niveau de boucles, leur position pouvant varier largement, mais la configuration et l'arrangement de la triade restent bien conservés.

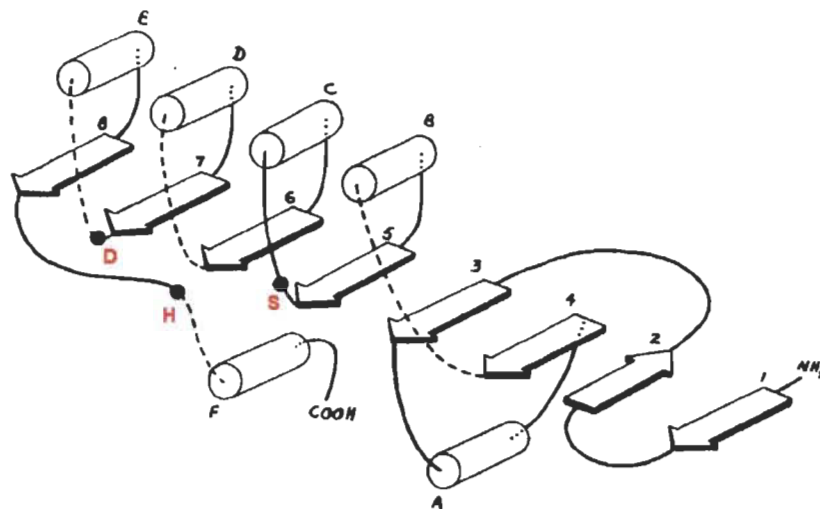


Figure 1.5 Topologie structurelle du repliement α/β hydrolase.

Le corps de la structure est formé par 8 brins bêta (flèches) formant un feuillet central entouré de 6 hélices alpha (représentées par des cylindres). Les résidus de la triade catalytique, Ser (S), Asp (D) et His (H) sont identifiés en rouge au niveau de leurs boucles respectives.

La figure 1.6 montre le repliement alpha/bêta hydrolase, les résidus de la triade catalytique se trouvant rapprochés par le repliement et placés dans leur configuration active (Ollis *et al.*, 1992; Nardini *et al.*, 1999). Ces résidus catalytiques se trouvent positionnés en configuration active au niveau d'une pochette qui permet à l'enzyme de lier le substrat et d'effectuer la catalyse. Différents types d'activités hydrolytiques sont identifiés pour les α/β hydrolases. Les α/β hydrolases à sérine catalysent des réactions hydrolytiques très variées et constituent le repliement protéique regroupant le plus de diversité fonctionnelle dans la nature. Le mécanisme de catalyse des α/β hydrolases à sérine fait intervenir une triade catalytique favorisant l'attaque d'une sérine nucléophile.

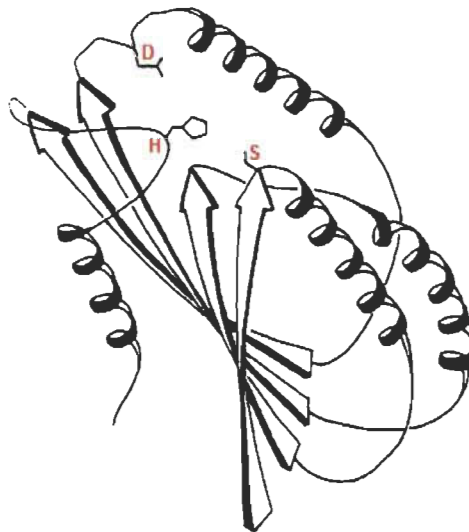


Figure 1.6 Repliement tridimensionnel des α/β hydrolases.

Ici le corps est simplifié et est représenté par un feuillet central formé de 4 brins bêta parallèles entouré de 4 hélices alpha. Les résidus de la triade catalytique sont montrés en rouge et en configuration active.

La triade consiste en une sérine positionnée entre le brin $\beta 5$ et l'hélice αC (le coude nucléophile), d'un résidu acide (l'aspartate ou le glutamate) situé dans une boucle après le brin $\beta 7$, et d'une histidine toujours située après le dernier brin bêta. Malgré le peu d'identité de séquence entre les protéines adoptant cette structure, et en dépit des importantes différences de fonction, la conservation du repliement ainsi que de la triade catalytique suggère que le repliement α/β constitue un échafaudage solide au niveau duquel différents types de fonctions peuvent évoluer.

1.4.3 Site actif

Le repliement α/β adopté par les carboxylesterases et des lipases forme une pochette de liaison du substrat au niveau de laquelle les résidus de la triade catalytique sont présentés dans une configuration active (E/D-H-S). Le mécanisme d'hydrolyse d'une liaison ester est représenté à la figure 1.7. Par la configuration géométrique de la triade et dans des conditions de pH favorable ($> pK_a$ His), l'aspartate chargé négativement interagit avec l'histidine (déprotoné) délocalisant les électrons au niveau de son cycle imidazole, ce qui a pour effet de permettre à l'azote (δ^1) du noyau imidazole d'agir comme une base générale et de déprotoner le groupe hydroxyle de la sérine. L'attaque nucléophile de la sérine se fait au niveau du carbocation du substrat et s'ensuit la formation d'un intermédiaire tétraédrique. Des résidus forment le trou de l'oxianion et stabilisent l'intermédiaire tétraédrique par la formation de ponts hydrogènes, lequel est décomposé par la libération du groupe alcool et la formation d'un acyl-enzyme. Une molécule d'eau effectue ensuite l'hydrolyse de l'acyl enzyme.

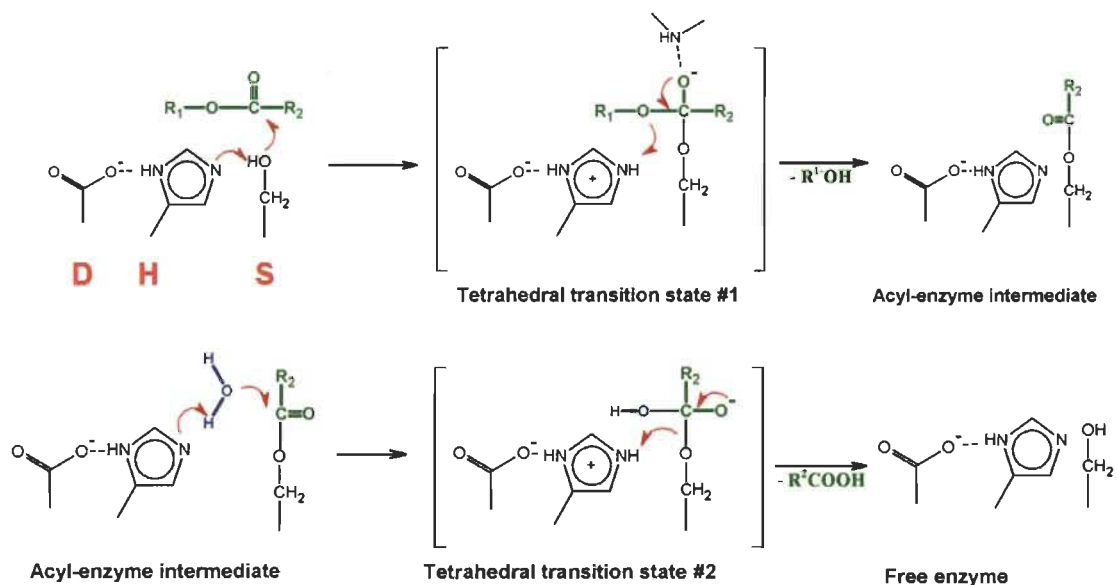


Figure 1.7 Mécanisme d'hydrolyse d'une liaison ester.

Les résidus actifs forment une triade catalytique (Ser-Asp/Glu-His) favorisant l'attaque nucléophile de la sérine catalytique. Il s'en suit l'hydrolyse de la liaison ester, le relargage d'un acide et la reformation de la triade active.

On assiste, ensuite, à la formation d'un deuxième intermédiaire tétraédrique lequel se décompose pour libérer le groupe acide et reformer la triade catalytique prête pour catalyser l'hydrolyse d'une nouvelle réaction. La figure 1.7 montre le mécanisme d'hydrolyse catalysée par les carboxylesterases et les lipases.

1.4.4 Spécificité et cinétiques enzymatiques

Une différence majeure existe entre les carboxylesterases et les lipases quant à leur spécificité et leur activité : les carboxylesterases (EC 3.1.1.1) sont des enzymes hydrosolubles qui catalysent l'hydrolyse ou la synthèse de liaisons esters au niveau de triglycérides ou autres molécules à courtes chaînes aliphatiques (de 2 à 10 carbones) semi-solubles en milieu aqueux; les lipases (EC 3.1.1.3) sont des enzymes hydrosolubles qui se spécialisent quant à elles dans l'hydrolyse et la synthèse de triglycérides à longues chaînes aliphatiques (> 10 carbones) et liposolubles. Les lipases sont inactives en milieu aqueux, et sont activées à l'interface lipide-eau, un phénomène appelé « activation interfaciale » (Verger *et al.*, 1997; Fojan *et al.*, 2000). Les carboxylesterases agissent suivant une cinétique de type Michaelis-Menten alors que les lipases montrent quant à elles une cinétique à tendance sigmoïdale due à l'activation interfaciale (Figure 1.8).

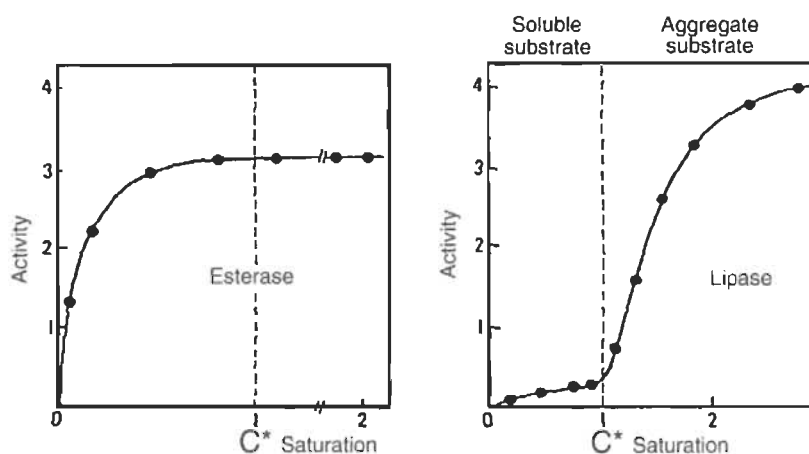


Figure 1.8 Différence de cinétique enzymatique entre les carboxylesterases et les lipases.

Les carboxylesterases montrent une cinétique de type Michaelis-Menten alors que les lipases sont activées par la présence d'interfaces.

En plus du corps α/β , où se trouve la triade catalytique, un domaine mobile supplémentaire appelé « lid » surplombe le site actif des lipases et forme un couvercle se refermant sur le site actif qui s'ouvre à la présence d'interfaces.

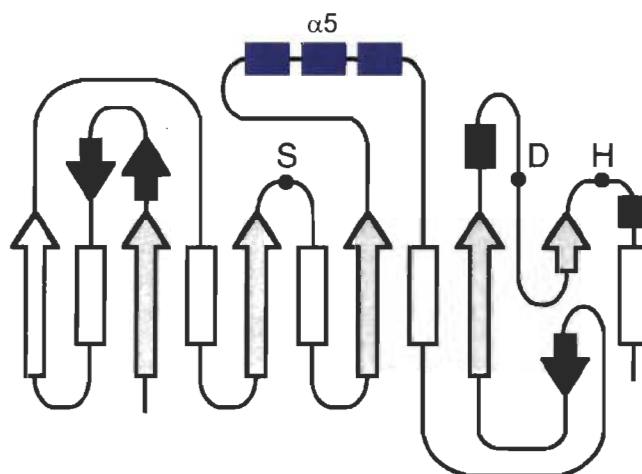


Figure 1.9 Topologie structurelle des lipases et éléments additionnels.
Lipase de *P. aeruginosa* (PDB no. 1EX9) *B. cepacia* (PDB no. 2LIP) et *P. cepacia* (PDB no. 1OIL). Les éléments de structures additionnels sont montrés en noir. Les hélices du « lid » sont montrées en bleu.

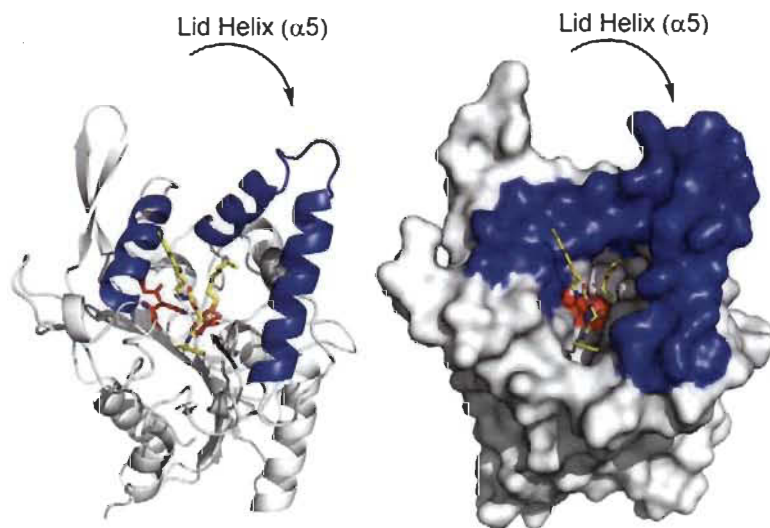


Figure 1.10 Structure du couvercle « lid ».
Structure cristalline de la lipase de *Pseudomonas cepacia* (PDB no. 1OIL). Le « lid » est représenté en bleu. Un inhibiteur (mimique de la tripalmitine), l'octyl-phosphonic acid 1,2-bis-octylcarbamoyloxy-ethyl ester, lié de façon covalente à la sérine catalytique en jaune.

Ce couvercle est normalement formé d'une ou plusieurs hélices, dont une particulière, mobile qui s'ouvre à l'interface lipide-eau, exposant le site actif au substrat où la catalyse peut donc s'opérer. La figure 1.9 montre la topologie d'une lipase de *Pseudomonas aeruginosa* associée au repliement alpha/bêta ainsi que l'insertion du domaine « *lid* » retrouvé chez les lipases qui est responsable de l'activation interfaciale. Les brins β et les hélices α « canoniques » sont représentés en blanc, alors que les insertions sont montrées en noir, les trois hélices formant le « *lid* » sont montrées en bleu (Figures 1.9 et 1.10). Les carboxylesterases, quant à elles, ne présentent normalement pas de domaine supplémentaire au-dessus du site actif, ce qui explique leur activité sur des composés solubles et l'absence d'activation interfaciale (Figure 1.11). Par contre, certaines carboxylesterases et nouvelles enzymes lipolytiques montrent un domaine supplémentaire similaire au « *lid* » appelé « *cap* ». Quoique présentant certaines similarités, le « *cap* » semble jouer un rôle dans la spécificité.

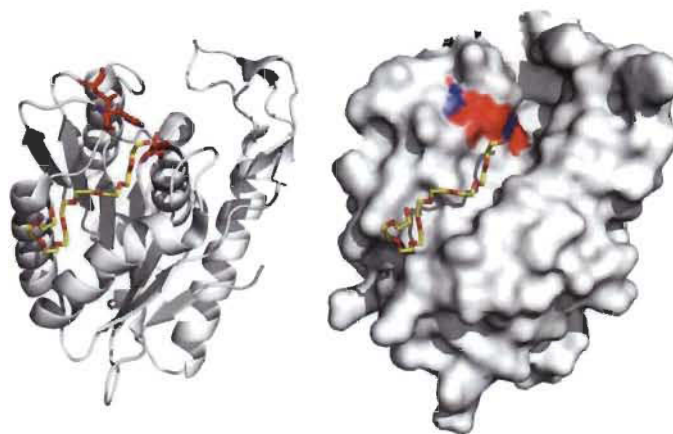


Figure 1.11 Structure des carboxylesterases.

Structure cristalline de la carboxylesterase de *Pseudomonas aeruginosa* PAO1 complexée avec le 2PE (PDB no. 3CN9). On note l'absence de « *cap* ». Les résidus de la triade catalytiques sont montrés en rouge.

1.4.5 Applications biotechnologiques des enzymes lipolytiques

Les enzymes lipolytiques sont utilisées dans de nombreux procédés biotechnologiques. Une liste regroupant les différentes applications industrielles des enzymes lipolytiques est présentée en annexe (Figure B.2). On note leur application dans

l'industrie des pâtes et papiers, dans la formulation de bioténergents ou dans la production de biodiesel (Gutierrez *et al.*, 2009). Les enzymes lipolytiques ont aussi la capacité de catalyser des réactions de transestérification en milieu organique. Vu leur énantio- et régiosélectivité, elles sont des catalyseurs intéressants pour la synthèse organique de certains médicaments, où il est difficile d'obtenir des composés optiquement purs par les méthodes de synthèse traditionnelles (Bornscheuer 2002). Dans l'industrie des pâtes et papiers, elles sont utilisées pour le contrôle des extractibles qui causent différents problèmes réduisant la qualité du papier. Les principaux composés estérifiés constituant la poix sont montrés dans la figure 1.12. (Prinsen *et al.*, 2012). (Gupta *et al.*, 2004; Sharma *et al.*, 2011; Jaeger et Eggert 2002; Sharma *et al.*, 2001; Jaeger et Reetz 1998; Jaeger et Reetz 1999; Shimada *et al.*, 1999; Bajaj *et al.*, 2010; Yu *et al.*, 2011).

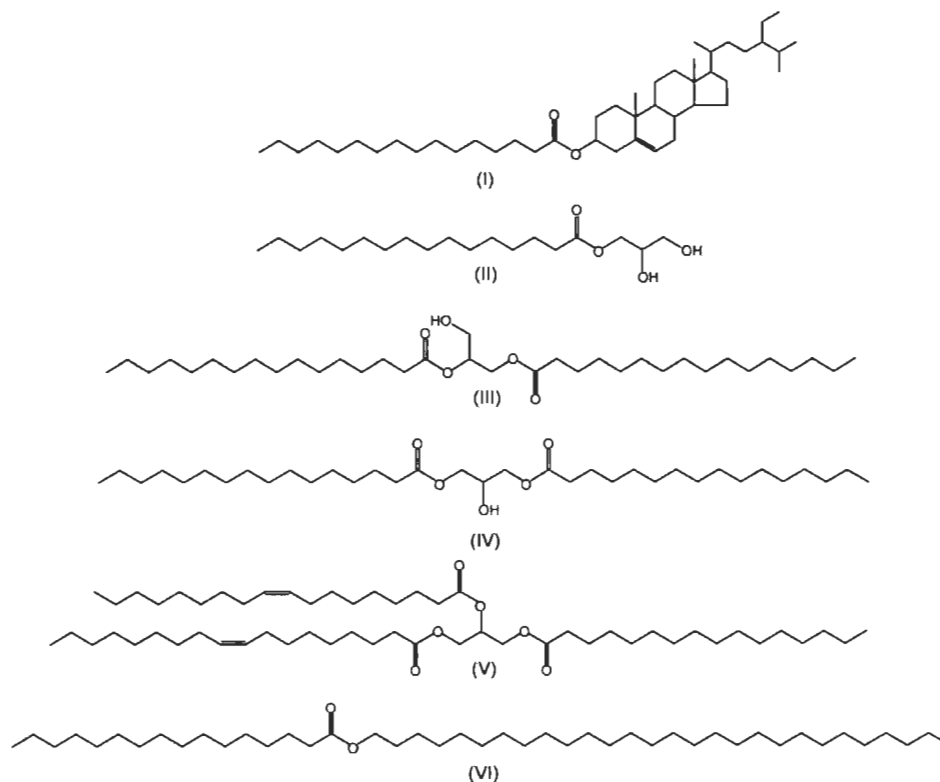


Figure 1.12 Principaux esters constituant les extractibles des tiges et du bois.

Analyse des lipides de la paille du blé. (I) sitosteryl palmitate, (II) 2,3 dihydroxypropyl hexadecanoate (1-monopalmitin), (III) 1,2 dipalmitin, (IV) 1,3 dipalmitin (V) dioloylpalmitin, (VI) Hexadecanoic acid octacosyl ester (del Riol *et al.*, 2013).

La réaction de synthèse de biodiesel catalysée par les lipases est montrée en figure 1.13.

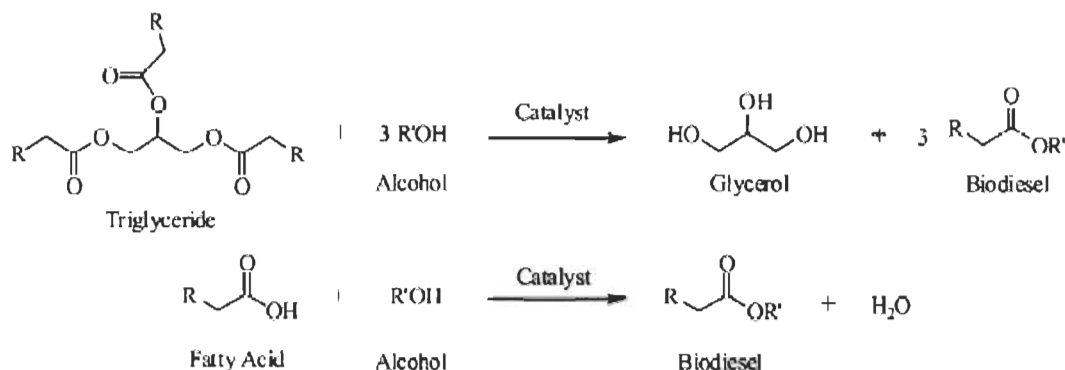


Figure 1.13 Réaction de synthèse de biodiesel catalysée par les lipases.

La transestérification entre les triglycérides ou les acides gras et le méthanol est utilisée pour la production de biodiesel.

1.4.6 Classification des enzymes lipolytiques

Les enzymes lipolytiques peuvent être classées en trois grands groupes sur la base de leur fonction et de leurs spécificités, les carboxylesterases (EC 3.1.1.1), les lipases (EC 3.1.1.3) et les phospholipases (EC 3.1.1.4). Les carboxylesterases et les lipases bactériennes ont été originellement classées en huit familles distinctes nommées (I-VIII) sur la base de leurs séquences et de leurs propriétés biochimiques (Arpigny et Jaeger, 1999). Dans cette classification, les « vraies lipases » (EC 3.1.1.3), c'est-à-dire procédant à l'activation interfaciale, sont classées dans la famille I qui se divise en six sous-familles. La famille II quant à elle regroupe des carboxylesterases et lipases ayant leur sérine catalytique dans un motif GDSL plutôt que dans le motif canonique habituel G-X-S-X-G. La famille IV ou HSL « hormone sensitive lipases » regroupe des lipases pour lesquelles un module de régulation supplémentaire est retrouvé en N-terminal et au centre de la séquence. Les familles III et V-VIII regroupent quant à elles des carboxylesterases d'origines mésophiles présentant différentes propriétés. Une classification plus exhaustive a été réalisée par Fisher et Pleiss, 2003, « The lipase engineering database » déposée sur le web à <http://www.led.uni-stuttgart.de/>. Cette classification regroupe de l'information sur les lipases, les estérases et autres alpha/bêta hydrolases. Plus de 800 protéines y sont associées à plus de 38 familles et

16 superfamilles. Similairement, « The microbial esterases and lipases database » regroupe plus de 800 estérases et lipases microbiennes classées en 15 familles de vraies estérases et 15 familles de lipases (Kang *et al.*, 2006). Cette classification recoupe certaines des familles initialement établies par Apigny et Jaeger, mais contient un éventail d'informations supplémentaires. Avec l'avènement de la métagénomique, plusieurs auteurs ont rapporté de nouvelles enzymes lipolytiques divergeant de la classification originale d'Apigny et Jaeger. Les familles IX-XIV ont donc été ajoutées à la classification originale (Rao *et al.*, 2013). Quoique de nature putative, certaines de ces nouvelles familles se recoupent. Des critères de classification exhaustifs devront être établis pour réconcilier ou distinguer ces nouvelles familles. Les enzymes lipolytiques thermostables synthétisées par des microorganismes thermophiles sont distribuées dans les différentes familles. Les microorganismes thermophiles sont une source de nouvelles enzymes thermostables intéressantes pour différentes applications biotechnologiques.

1.5 Les microorganismes thermophiles : une source d'enzymes thermostables

Les microorganismes thermophiles sont pour la plupart des organismes unicellulaires se développant à haute température (45-122 °C) et sont une source importante d'enzymes thermostables (Madigan *et al.*, 2006, Takai *et al.*, 2008). Les microorganismes dits proprement thermophiles se développent optimalement entre 50 et 80 °C alors que les microorganismes hyperthermophiles ont des températures optimales de croissance supérieures à 80 °C. Ces microorganismes se développent dans différentes niches écologiques de la planète et sont des organismes robustes qui montrent un fort potentiel pour différentes applications industrielles. La comparaison de la structure de différentes enzymes thermostables avec celles de leurs homologues mésostables constitue l'approche classique pour élucider certains éléments structuraux potentiellement thermostabilisants chez les versions thermophiles. L'introduction de ces éléments chez les versions mésophiles ou encore l'annihilation de ces éléments chez les versions thermophiles peuvent permettre d'élucider le rôle de certains éléments structuraux dans la thermostabilité. Le succès pour l'identification d'éléments stabilisants par cette méthode repose souvent sur la haute conservation du repliement

entre les deux protéines étudiées. Les microorganismes se divisent en deux classes, modérément thermophiles et hyperthermophiles (Szilagyi et Zavodszky, 2000). Les plus importants représentant des microorganismes thermophiles sont les bactéries et les archaea (Stetter 2006).

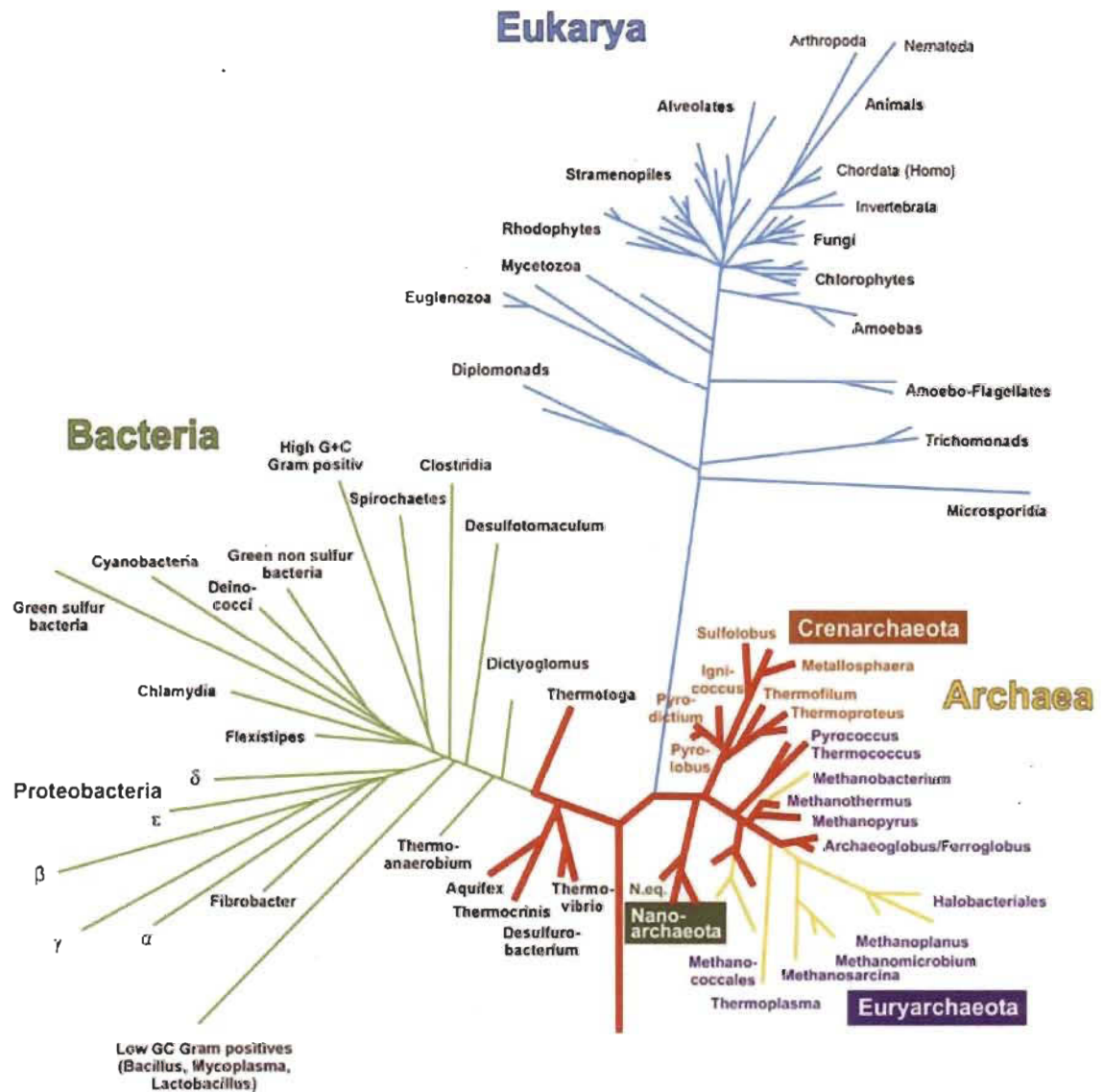


Figure 1.14 Phylogénie du vivant.

Arbre phylogénétique représentant les trois règnes du vivant basé sur la petite sous-unité ribosomale (ARNr). Les branches en rouge près de la racine représentent les microorganismes hyperthermophiles (Stetter 2006).

Une analyse phylogénétique basée sur la petite sous-unité ribosomale (ARNr 16S chez les procaryotes) offre un portrait intéressant. Les microorganismes thermophiles et hyperthermophiles se retrouvent près de la racine, ce qui suggère, dans le cas où celle-ci constitue l'origine, que la vie a pu apparaître à haute température, telle que celles retrouvées originellement pour ensuite s'adapter aux différentes températures, telles que celles retrouvées à la surface de la Terre aujourd'hui, par évolution darwinienne (Figure 1.14). Un débat est toujours ouvert à savoir si celles-ci se sont adaptées à haute température à partir de descendants mésophiles ou si c'est plutôt le cas inverse. Il ne faut pas oublier que ce processus peut être bidirectionnel. Différentes études ont démontré que les enzymes thermostables et psychrostable ont un meilleur potentiel évolutif (Bloom 2006). On retrouve des hyperthermophiles chez les archaea et chez les bactéries. Les microorganismes hyperthermophiles sont représentés par les branches en rouge, près de la racine (Figure 1.14). Le genre bactérien *Thermotoga* est un exemple de bactéries hyperthermophiles qui se développe entre 55 et 90 °C avec une température idéale de croissance à 80 °C. Les branches en vert près de celles en rouge représentent les bactéries thermophiles. On note les *Dietyoglomus*, *Thermoanaerobium* et les bactéries Gram positives à faible proportion en G + C tels les *Bacillus*. Ces microorganismes sont sujets à synthétiser des protéines thermorésistantes et des enzymes thermostables intéressantes pour l'industrie.

1.6 Repliement et stabilité des protéines

Le processus de dénaturation des protéines a été découvert par Chick et Martin en 1910. Le fait que la dénaturation était liée à la perte de la structure tridimensionnelle d'une protéine par un processus de « dépliement » a été avancé par Hsien Wu en 1929 (Edsall 1995), alors que Tanford propose en 1968 une théorie pour la dénaturation des protéines. Ensuite, Pace formula en 1975 une description mathématique de l'état natif et dénaturé des protéines. Historiquement, puisque les acides et les bases ont été les premiers dénaturants connus des protéines, il a été immédiatement suggéré que les interactions électrostatiques puissent jouer un rôle prédominant dans le repliement des protéines (Linderstrøm-Lang 1949; Perutz 1963). Il y a deux façons par lesquelles les

interactions électrostatiques peuvent affecter la stabilité d'une protéine. Les effets électrostatiques peuvent être stabilisants ou déstabilisants, ceux-ci sont décrits par le modèle de Poisson-Boltzman (Warshel et Papazyan 1998; Michael Gilson, 1995; Fogolari *et al.*, 2002). Des répulsions électrostatiques surviennent quand une protéine possède un important contenu en groupements ionisables. Comme la charge nette d'une protéine augmente en condition très acide (charge positive) ou très basique (charge négative), les répulsions électrostatiques non spécifiques augmentent aussi pouvant déstabiliser la structure native d'une protéine, ce qui explique la dénaturation par le pH.

La seconde façon de concevoir l'effet des interactions électrostatiques sur la structure et la stabilité d'une protéine consiste en ses interactions spécifiques entre groupements ionisables de charges opposées. Par exemple, le pairage ionique (ponts salins) survient quand deux résidus de charges opposées se retrouvent à proximité et dans une orientation favorable. Contrairement aux effets déstabilisants prévus par les mécanismes de répulsion électrostatique, les ponts salins peuvent stabiliser la structure. Il est clair que les ponts salins sont importants pour la stabilité des protéines, mais plusieurs évidences suggèrent que ceux-ci ne constituent pas la principale force pour le repliement (Pace et Grimsley, 1988). Une des principales évidences suggérant que les ponts salins ne sont pas une force dominante pour le repliement des protéines globulaires en solution est que ceux-ci ne sont pas toujours conservés durant l'évolution et que la plupart des protéines (mésostables) ne montrent qu'une faible proportion de ce type d'interaction. Plusieurs protéines dont les structures cristallines sont connues ne montrent aucune interaction entre groupements de charges opposées. De plus, plusieurs études ont démontré qu'en général, la contribution de la composante électrostatique d'un pont salin sur l'énergie libre du repliement est très modeste, environ 1-3 kcal/mol, ce qui est environ 5 à 10 fois inférieur aux valeurs observées pour les interactions hydrophobes.

Quelles sont les forces qui régissent la thermostabilité des protéines? Est-ce que ces forces peuvent aussi régir le repliement? En 1936, Mirsky et Pauling (Mirsky et Pauling, 1936) conclurent que les ponts hydrogènes (ponts-H) sont une force majeure dans la stabilité des protéines. Ce point de vue a été renforcé dans les années 50 quand

Pauling démontra que l'hélice alpha et le feuillet bêta étaient des éléments structuraux stabilisés par des ponts hydrogènes, et que ceux-ci étaient aussi à la base de la formation et des interactions entre les éléments de structures secondaires. En 1959, Walter Kauzmann suggère que la principale force déterminant le repliement des protéines en solution vient de la forte aversion des groupements non polaires pour l'eau du milieu aqueux (Kauzmann 1959). Il démontre aussi que la plupart des protéines sont dénaturées en milieu non polaire. Ses idées furent avant-gardistes puisque peu d'informations sur la structure des protéines, d'analyse par mutagenèse et de mesure de stabilité étaient connues à l'époque. Ces observations suggèrent que les effets hydrophobes constituent la principale force impliquée dans le repliement des protéines solubles et dans leur stabilité en solution, les autres types de forces étant par ailleurs impliquées dans la stabilité de l'état replié, natif (Pace 1996; Pace et Grimsley 2011).

D'autres études ont par la suite renforcé l'idée que les effets hydrophobes doivent être une force dominante du repliement. La dénaturation des protéines par différents solvants organiques suggère que les ponts hydrogènes ne sont pas non plus une force dominant le repliement des protéines en solution (Singer 1962). Une évidence particulière de l'importance des effets hydrophobes sur le repliement des protéines est issue de la comparaison de l'effet de différents alcools sur la stabilité des protéines. Plusieurs résultats démontrent que l'éthanol et le propanol sont beaucoup moins déstabilisants que les glycols correspondant (éthylène glycol et propylène glycol), ce qui suggère fortement que les ponts hydrogènes sont moins importants que les effets hydrophobes, possiblement la force dominante pour la stabilité des protéines en solution aqueuse.

Outre les effets hydrophobes, les ponts hydrogènes, les interactions ioniques, les ponts disulfures et les effets induits et forces de van der Waals constituent les principales forces connues impliquées dans la stabilité des protéines (Dill 1985, 1990, 2008; Pace 1996). Ces différentes interactions sont représentées dans la figure 1.15.

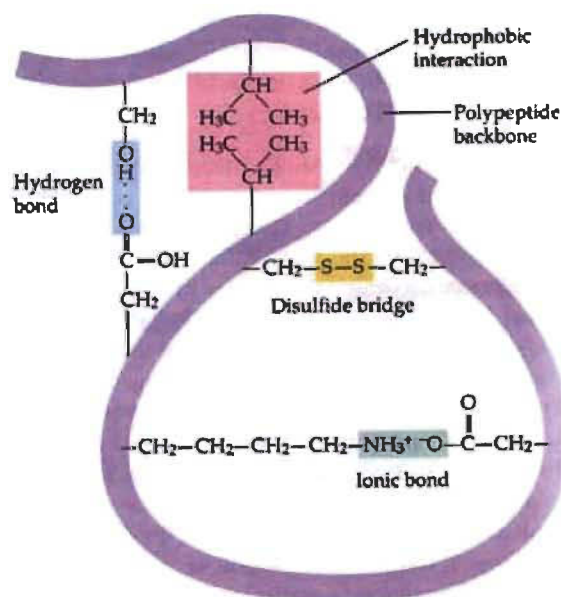


Figure 1.15 Principaux déterminants structuraux de la stabilité des protéines.

Principales interactions entre les chaînes latérales des acides aminés constituant les protéines impliquées dans leur stabilité conformationnelle.

1.6.1 Détermination de la stabilité conformationnelle des protéines

Afin de mesurer expérimentalement la stabilité conformationnelle d'une protéine, il est nécessaire de déterminer la constante d'équilibre de dénaturation (K) ainsi que le changement d'énergie libre (ΔG) entre l'état natif et dénaturé ($N \leftrightarrow D$). La proportion de molécules de protéines natives et dénaturées en solution ainsi que les paramètres thermodynamiques associés au processus de dénaturation peuvent être déterminés expérimentalement par différentes méthodes spectroscopiques. Par exemple, le spectre de dichroïsme circulaire (CD) en UV lointain est caractéristique des structures secondaires, structures en hélices alpha et en feuillet bêta (Figure 1.16). Le signal CD à 222 nm est par exemple une sonde conformationnelle pratique pour suivre le niveau d'hélicité durant le processus de dénaturation d'une protéine en fonction d'une condition dénaturante comme la température. Le spectre de dichroïsme circulaire en UV rapproché est quant à lui caractéristique de la mobilité des résidus aromatiques et peut être corrélé aux changements de la structure tertiaire. Le signal CD à 290 nm correspond par exemple à la mobilité de résidus tryptophane normalement peu mobile lorsqu'enfoui dans la structure native d'une protéine.

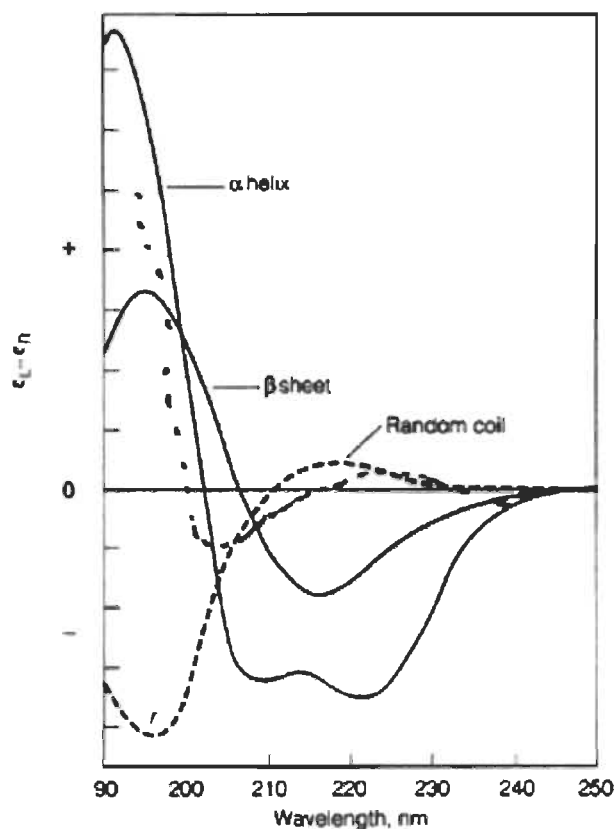


Figure 1.16 Spectres de dichroïsme circulaire en UV lointain.

Les liaisons amides dans les protéines absorbent la lumière polarisée circulairement de différentes façons dépendamment de leur configuration. Les liaisons amides impliquées dans la formation d'hélices ont un spectre d'absorption différent de celles impliquées dans la formation de feuillet bêta ou de structures désordonnées. La figure montre les spectres d'absorption (différence d'absorption de lumière polarisée circulairement gauche et droite) caractéristiques des différentes structures secondaires.

Les hélices alpha montrent un spectre de dichroïsme circulaire particulier présentant deux minima soit à 208 et 222 nm alors que les structures en feuillets bêta montrent un seul minimum à 215 nm. Les structures désordonnées montrent quant à elles un minimum d'absorption à 205 nm. La mesure du signal de dichroïsme circulaire à 222 nm en fonction de la température permet de mesurer le changement d'hélicité au cours de la dénaturation thermique d'une protéine et donc de mesurer la transition entre l'état natif et l'état dénaturé. Cette transition peut être utilisée pour calculer les paramètres thermodynamiques associés à la dénaturation thermique.

Le changement de fluorescence intrinsèque associé à l'exposition relative de résidus aromatiques initialement enfouis dans la structure globulaire (relaxation par le solvant) constitue aussi une sonde efficace pour le suivi de la dénaturation conformationnelle d'une protéine. Le déplacement de la longueur d'onde d'émission maximale de fluorescence (λ_{\max}) en fonction de la température est utilisé pour mesurer les changements conformationnels de la structure tertiaire d'une protéine.

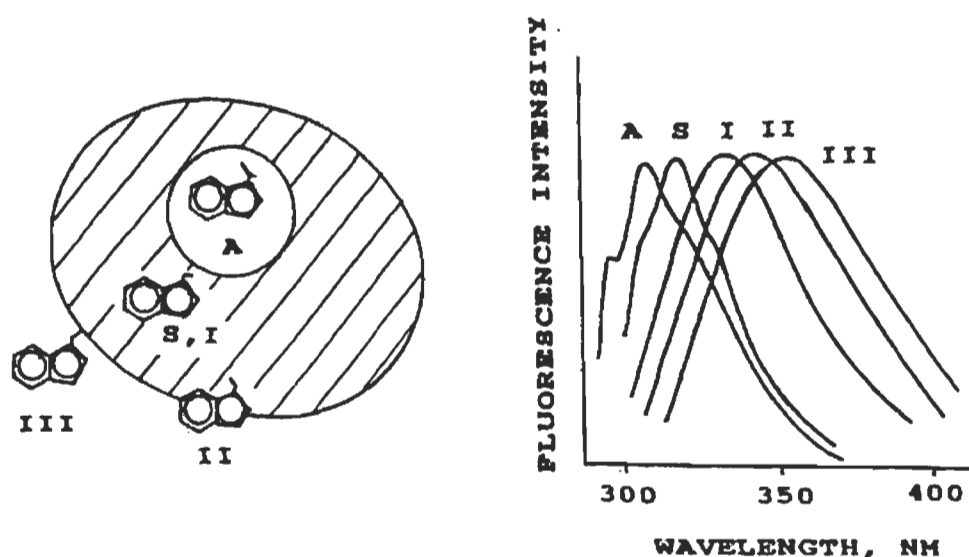


Figure 1.17 Relaxation par le solvant.

Un résidu aromatique comme un tryptophane, initialement enfoui à l'intérieur du cœur hydrophobe d'une protéine a une longueur d'onde d'émission maximale de fluorescence autour de 320-330 nm (S) alors que ce même résidu tryptophane émet maximalement autour de 350 nm lorsque complètement exposé au solvant aqueux (III).

En assumant que le processus de dénaturation s'opère entre deux états (natif et dénaturé), entre la fraction de molécules natives (f_N) et dénaturées (f_D) et que $f_N + f_D = 1$, la valeur mesurée du signal (y) à n'importe quel point de la courbe de dénaturation est donnée par :

$$y = y_N f_N + y_D f_D$$

Les valeurs de y_N et y_D représentent les valeurs de signal (y) de l'état natif et dénaturé respectivement. En combinant ces équations, on obtient :

$$f_D = (y_N - y) / (y_N - y_D)$$

La constante d'équilibre (K) peut se calculer comme suit :

$$K = f_D / (1 - f_D) = f_D / f_N = (y_N - y) / (y - y_D)$$

Le changement d'énergie libre (ΔG) peut quant à lui se calculer par l'équation suivante :

$$\Delta G = -RT \ln K = -RT \ln (y_N - y) / (y - y_D)$$

On peut ensuite déterminer la valeur de la température de fusion (T_m), qui correspond à la température à laquelle la moitié des molécules de protéines sont dans l'état natif et dans l'état dénaturé ($f_N = f_D$). Cette valeur peut être déterminée en portant en graphique la relation entre l'énergie libre (ΔG) et la température dont l'équation de la droite se décrit par :

$$\Delta G = \Delta H - T\Delta S$$

Le T_m correspond à la température où le changement d'énergie libre est nul, T à laquelle $\Delta G = 0$. La différence d'enthalpie au T_m peut donc se calculer comme suit :

$$\Delta H_m = T_m \Delta S$$

Nous voulons maintenant déterminer la stabilité conformationnelle à la température standard de 25 °C. Cette opération requiert d'extrapoler la mesure de l'énergie libre de la mince plage de températures auxquelles le dépliement se produit à la température ambiante de 25 °C. L'équation de van't Hoff permet de mesurer le changement d'enthalpie (ΔH).

$$d(\ln K) / d(1/T) = -\Delta H/R$$

La relation de van't Hoff ($\ln K$ en fonction de $1/T$) n'est pas linéaire, puisque le ΔH varie en fonction de la température, ce qui indique que la capacité calorifique du réactant (protéine à l'état natif) et du produit (protéine à l'état dénaturé) diffèrent.

$$d(\Delta H)/d(T) = C_p(D) - C_p(N) = \Delta C_p$$

En déterminant le T_m , l'enthalpie (ΔH) au T_m (ΔH_m) et la différence de capacité calorifique (ΔC_p), on peut calculer ΔG , à n'importe quelle température selon :

$$\Delta G(T) = \Delta H_m (1 - T/T_m) - \Delta C_p [(T_m - T) + T \ln (T/T_m)].$$

On réfère à la valeur de ΔG à 25 °C ou en absence de dénaturant comme mesure de la stabilité conformationnelle d'une protéine. La différence d'énergie libre $\Delta\Delta G$ à 25 °C entre deux protéines distantes ou encore entre le type sauvage et une version mutante est prise comme mesure pour la comparaison de leur stabilité conformationnelle et donc pour mesurer l'impact de mutations particulières sur la stabilité (Pace *et al.*, 1989).

1.6.2 Déterminants structuraux de la thermostabilité des protéines

Les principaux facteurs responsables de conférer la thermostabilité des enzymes thermo- et hyperthermostables ont été revisités par Kumar (Kumar *et al.*, 2000). Les éléments responsables du repliement des protéines en solution, tels que les résidus responsables de l'empaquetage du cœur hydrophobe, de la formation de ponts hydrogènes ainsi que des forces de van der Waals sont stabilisants. Mais quels sont les déterminants de la thermostabilité? La comparaison de la structure d'enzymes synthétisées par des microorganismes thermophiles avec leurs homologues mésophiles a permis de déterminer certains éléments structuraux responsables de conférer la thermostabilité chez les types thermophiles. Par exemple, une analyse comparative de la structure de 18 familles de protéines thermostables et mésostables a permis d'établir que l'hydrophobicité, le niveau d'empaquetage du cœur hydrophobe, la composition en résidus polaires et non polaires en surface, la présence de ponts disulfures et les

substitutions de prolines étaient très similaires entre les versions thermophiles et mésophiles étudiées. Ces caractéristiques ont souvent été attribuées comme facteurs responsables de l'importante stabilité de certaines protéines thermophiles par le passé. Par ailleurs, le nombre de ponts hydrogènes entre les chaînes latérales ainsi que le nombre de ponts salins augmente de façon significative dans les structures thermophiles. De plus, une présence accrue de résidus d'arginine et de tyrosine concomitante avec une réduction du nombre de cystéine et sérine a été observée chez les enzymes thermostables (Kumar *et al.*, 2000). La comparaison de la structure de l'indolglycerol phosphate synthase de *Thermotoga maritima* avec celle de son homologue de chez *E. coli* et de chez *Sulfolobus solfaraticus* ainsi que la comparaison de la version thermostable de l'adénylate kinase de *Geobacillus stearothermophilus* avec celle de son homologue mésophile *Bacillus subtilis* et psychrophile *Bacillus globisporus* montre une nette corrélation entre le nombre de ponts salins et la thermostabilité. Similairement, un accroissement de la teneur en ponts salins est directement relié à l'accroissement de la thermostabilité entre la version psychrostable, mésostable et thermostable de l' α -amylase de *Bacillus amyloliquefaciens*, *Bacillus stearothermophilus* et *Bacillus licheniformis* respectivement (Knöchel *et al.*, 2002; Bae et Phillips, 2004; Tomazic et Klivanov 1988).

1.6.3 Enzymes lipolytiques et thermostabilité

Les enzymes lipolytiques sont distribuées dans tous les règnes du vivant et sont bien représentées chez les microorganismes thermophiles et hyperthermophiles (Levinson *et al.*, 2009). L'étude de la thermostabilité des enzymes lipolytiques, à l'heure où ces enzymes occupent la principale part du marché des enzymes industrielles, fait l'objet d'études intensives. Récemment, différentes études ont porté sur la détermination d'éléments structuraux impliqués dans la thermostabilité des carboxylesterases et des lipases du genre *Bacillus*. La comparaison de la structure de la Brefeldine A (BFAE) de *Bacillus subtilis* (mésostable) avec la carboxylesterase EST2 d'*Alicyclobacillus acidocaldarius* (thermostable) et de la carboxylesterase AFEST d'*Archaeoglobus fulgidus* (hyperthermostable) ainsi que l'étude des déterminants structuraux révèle que

les ponts salins jouent un rôle majeur dans la thermostabilité (Del Vicchio *et al.*, 2002; Mandrich *et al.*, 2006; Pezzullo *et al.*, 2013). La comparaison structurale de la phosphotriesterase SsoPox de *Sulfolobus solfaraticus* avec ses homologues mésostables de *Pseudomonas diminuta* et *Agrobacterium radiobacter* a montré que la présence accrue de ponts salins souvent formés en réseaux sont un important déterminant de la thermostabilité de SsoPox (Del Vecchio *et al.*, 2009). Les enzymes lipolytiques thermostables sont aussi connues pour montrer une résistance accrue à différents solvants organiques tels que les alcools et différents solvants non polaires, ce qui peut s'expliquer par la présence accrue de ponts salins dans ce type d'enzymes.

1.6.4 Les ponts salins : interactions clés chez les protéines thermophiles

Les interactions électrostatiques jouent un rôle essentiel dans la structure et la fonction biologique des protéines (Perutz 1978, Warshel 1981). Elles sont à la base du pouvoir catalytique des enzymes et jouent un rôle important dans leur stabilité. Les interactions électrostatiques sont la clé pour le fonctionnement des enzymes et l'évolution des microorganismes thermo- et hyperthermophiles (Vertriani 1998), mais comment ces interactions stabilisent-elles les protéines de ces organismes thermophiles? Les interactions électrostatiques peuvent être déstabilisantes dues à des phénomènes de répulsion ou stabilisantes par la formation de ponts salins. Les ponts salins sont un type d'interaction spécifique formée entre le groupe amine d'une chaîne latérale basique (His, Lys ou Arg) et le groupe carboxylique d'une chaîne latérale acide (Asp ou Glu) quand chacun de ceux-ci porte une charge opposée (positive pour les résidus basiques et négative pour les résidus acides) et que ceux-ci sont dans une orientation propice et que le centre de masse de l'azote du groupe amine (NH_3^+) du résidu basique impliqué est situé à moins de 4 Å de celui de l'oxygène du carbonyle du groupe carboxylate (COO^-). La formation d'un pont salin dépend donc du pH et de l'identité des chaînes latérales impliquées. En effet, l'identité des résidus impliqués, leur niveau d'enfouissement et de leur pKa font que la formation d'un pont salin dépend du pH.

Par exemple, la chaîne latérale de l'histidine exposée en surface d'une protéine a normalement un pKa modéré (~ 6.1) contrairement à la chaîne latérale de l'arginine qui a un pKa beaucoup plus élevé (~ 12.4). Un pont salin entre un résidu aspartate (pKa ~ 3.7) et un résidu arginine en surface peut se former à l'intérieur d'une plage de pH beaucoup plus large qu'un pont entre un résidu aspartate et un résidu histidine par exemple. Les ponts salins s'avèrent un type d'interaction qui semble être directement lié à la thermophilicité des protéines et donc des éléments cibles potentiels en vue de modifier par une approche rationnelle la stabilité ou la réponse au pH d'un biocatalyseur.

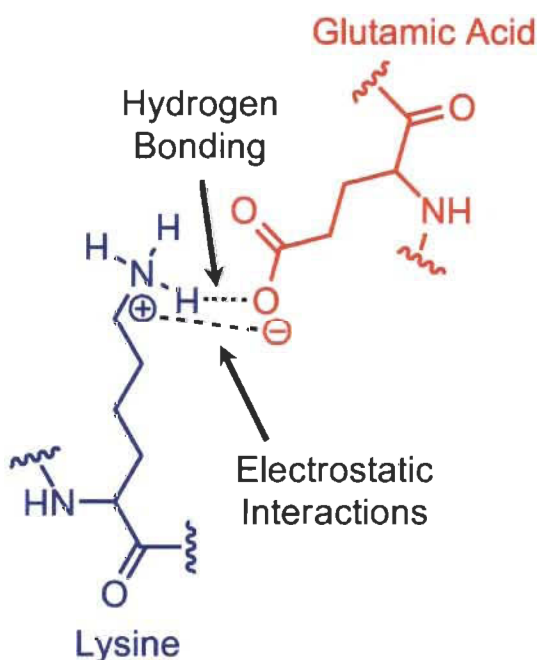


Figure 1.18 Formation d'un pont salin.

Exemple de la formation d'un pont salin entre un glutamate et une lysine.

Les interactions électrostatiques sont largement considérées dans les différentes approches de design de protéines (Marshall *et al.*, 2002; Vizcarra et Mayo, 2005). Contrairement aux interactions hydrophobes, les interactions électrostatiques semblent tirer leur efficacité à haute température (Elcock 1998). Quoique les interactions électrostatiques soient stabilisantes dans la structure d'une protéine pliée, à haute température, leur implication possible dans le repliement reste encore mal comprise.

1.6.5 Les interactions électrostatiques peuvent-elles jouer un rôle dans le repliement de certaines protéines?

À température modérée, les effets hydrophobes sont connus comme résultant d'un phénomène de nature principalement entropique qui provient de l'incapacité de formation de liaisons hydrogènes, hautement dynamiques, entre les molécules d'eau et le solvant non polaire (Kauzmann 1959). Par contre, à haute température, les molécules d'eau deviennent plus mobiles et peuvent donc pénétrer davantage à l'intérieur du cœur hydrophobe d'une protéine, réduisant le gain énergétique des effets hydrophobes, ceux-ci étant compensés par un effet enthalpique (Baldwin 1986), ce qui fait que les effets hydrophobes sont modérément affectés par l'accroissement de la température, mais par contre moins efficaces à basse température expliquant la dénaturation par le froid de plusieurs protéines. Il semble donc que les interactions hydrophobes tirent leur pleine efficacité à température modérée, ce qui explique leur rôle dominant dans le repliement et dans la stabilité des protéines mésostables. Cette observation suggère que la présence accrue de ponts salins dans les protéines thermostables contribue à un apport stabilisant supplémentaire à haute température, où la compensation par effet enthalpique atteint possiblement un plateau pour le gain énergétique des effets hydrophobes. Les effets hydrophobes gouvernent probablement le repliement et la stabilité des protéines mésostables, à température ambiante ou modérée, mais quels sont les déterminants du repliement et de la stabilité des protéines psychostables et thermostables, voire hyperthermostables?

Il est raisonnable de penser que les interactions hydrophobes doivent jouer un rôle moins important dans la thermostabilité des enzymes thermo- et hyperthermostables qui opèrent à haute température. L'eau, hautement polarisable, peut interférer par la formation de ponts hydrogènes compétitifs avec ceux impliqués dans la stabilisation des éléments de structures des protéines. Il est connu que les ponts hydrogènes et les ponts salins enfouis soient davantage stabilisant que les ponts en surface (Tisot *et al.*, 1996; Zhou *et al.*, 2002) quoiqu'un pont salin enfoui peut être moins stabilisant qu'une interaction hydrophobe d'empaquetage similaire, dû au coût de désolvatation et d'enfouissement des résidus polaires (Waldburger et Sauer; Hendsch et Tidor 1993). Les

enzymes lipolytiques thermo- et hyperthermostables montrent un nombre important de ponts salins, sont stables en milieu organique et à haute température. Un important contenu en ponts salins montre un lien direct avec la thermostabilité, mais est-ce plutôt les forces de répulsion électrostatique entre les résidus de même charge (haut contenu en résidus chargés), combiné aux effets hydrophobes qui peuvent être déterminants pour le repliement, les ponts salins formés dans la structure finale contribuant seulement à la stabilité de l'état natif? Il ne faut pas exclure le rôle des autres interactions comme les pont-H et l'empaquetage des chaînes latérales de ces mêmes résidus chargés.

1.7 Problème du repliement des protéines

Un des plus importants problèmes fondamentaux encore non résolus en biologie moléculaire moderne demeure celui du repliement des protéines. Quelle est la relation qui existe entre l'information à une dimension contenue dans la séquence d'acides aminés d'une protéine et l'information à trois dimensions formée par sa structure tridimensionnelle? Ce problème comporte différentes énigmes qui peuvent se résumer par les trois questions suivantes : (a) Peut-on prédire la structure d'une protéine sur la base de la seule connaissance de sa séquence? (b) Existe-t-il un code pour le repliement des protéines? (c) Quels sont les mécanismes par lesquels une séquence d'acides aminés adopte sa structure finale aussi rapidement? (Levinthal 1968, Behe *et al.*, 1991; Zwanzig *et al.*, 1992; Lattman *et al.*, 1993; Dill 2008, 2012 et Shaeffer *et al.*, 2011).



Figure 1.19 Repliement cotraductionnel des protéines.

Une séquence d'ARNm est traduite en séquence d'acides aminés par le ribosome. Le polypeptide émergent commence à se replier au fur et à mesure que sa chaîne polypeptidique s'allonge.

1.7.1 Solution au problème du repliement : code ou énergie?

Est-ce que la structure native d'une protéine, unique et stable, est déterminée uniquement par sa séquence? Si oui, est-il possible de résoudre le problème du repliement? L'hypothèse thermodynamique, ou le dogme d'Anfinsen stipule que seule la séquence détermine le repliement d'une protéine (Anfinsen 1973). Les résultats des expériences d'Anson et Mirsky (1931) montrent que le repliement de l'hémoglobine est un processus réversible. Par la suite, il a été démontré que le repliement de plusieurs autres protéines purifiées était aussi réversible *in vitro* suite leur dénaturation suggérant que seule la séquence détermine le repliement (Nojima *et al.*, 1977). Par contre, d'autres protéines ne possèdent pas ces mêmes propriétés de renaturation et certaines peuvent les présenter, mais seulement en présence de molécules accessoires, tel que les protéines chaperonnes. Les interactions protéines-protéines sont-elles à la base du repliement de certaines protéines? Mais est-ce que seule la séquence détermine le repliement et l'atteinte de la structure finale? Si tel est le cas, est-ce que le repliement est un processus opérant localement, au fur et à mesure que la protéine synthétisée émerge du ribosome? Ou, est-ce que le mécanisme du repliement est un processus se réorganisant jusqu'à l'atteinte d'un minimum énergétique thermodynamiquement stable? Malgré l'asymétrie de la structure d'une protéine, celle-ci peut dans des conditions particulières, former un cristal permettant l'élucidation de sa structure par cristallographie aux rayons X, ce qui suggère que la séquence d'une protéine acquière une structure unique. Cette unicité suggère par le fait même que la structure formée adopte un minimum énergétique, mais comment la protéine acquiert cette conformation unique si rapidement?

Le paradoxe de Levinthal suggère l'expérience de pensée suivante : si la séquence d'acides aminés d'une protéine explore l'ensemble des possibilités conformationnelles permises par sa chaîne polypeptidique avant l'atteinte de sa structure finale, vu le très grand nombre de degrés de liberté permis pour le repliement d'une séquence d'acides aminés de longueur standard, une estimation astronomique de 3^{300} ou 10^{143} conformations possibles pouvant être explorées par une protéine de longueur standard avant l'atteinte de sa structure finale, optimalement stable, ce qui nécessiterait un temps supérieur à l'âge de l'univers. Ce constat absurde soulève le fait que le

processus du repliement des protéines doit s'effectuer selon un chemin préférentiel, un chemin ou « paysage énergétique » définit plutôt que par un échantillonnage des espaces conformationnels, puisque l'on sait qu'une protéine se replie presque instantanément, dans l'ordre des millisecondes, voire des microsecondes dans certains cas (Plotkin *et al.*, 2002, Bryngelson *et al.*, 2004).

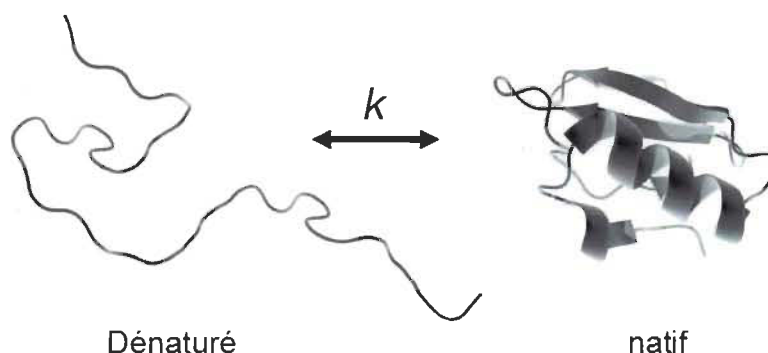


Figure 1.20 Repliement et dénaturation réversible des protéines.

Certaines protéines peuvent se renaturer par elles-mêmes après dénaturation, ce qui suggère que l'information pour le repliement est contenue dans la séquence.

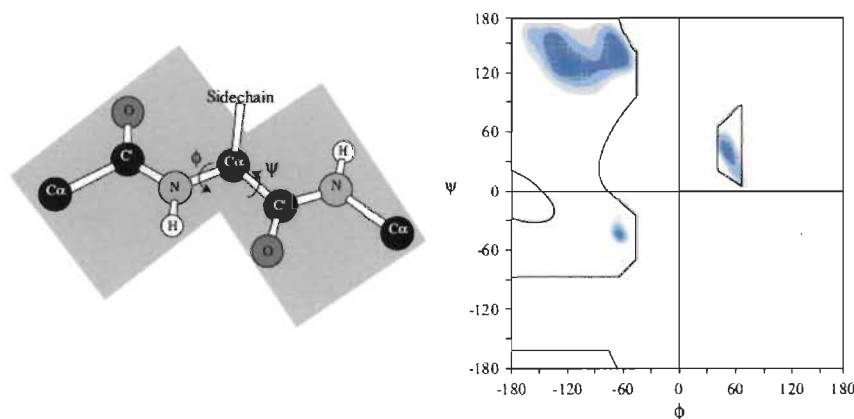


Figure 1.21 L'unité peptidique et les degrés de liberté permis.

Assumant une complète indépendance des deux angles de torsion ϕ et ψ , il y a trois configurations possibles par torsion, $3 \times 3 = 9$ configurations par unité peptidique donc $9^{100} = 10^{95}$ conformères possibles pour une simple protéine de 100 acides aminés (Rose *et al.*, 2006).

Différentes théories pour décrire le repliement des protéines ont vu le jour, passant de chemins de repliement (*folding pathways*) aux paysages énergétiques (*energy landscapes*) à la théorie des entonnoirs (*funnels*) (Leopold *et al.*, 1992; Onuchic *et al.*, 1994; Dill 1997). Existe-t-il un code pour le repliement des protéines? La deuxième partie du code génétique consiste à comprendre comment est encodée l'information structurale d'une protéine dans sa séquence d'acides aminés, donc dans la séquence d'acides nucléiques de son gène.

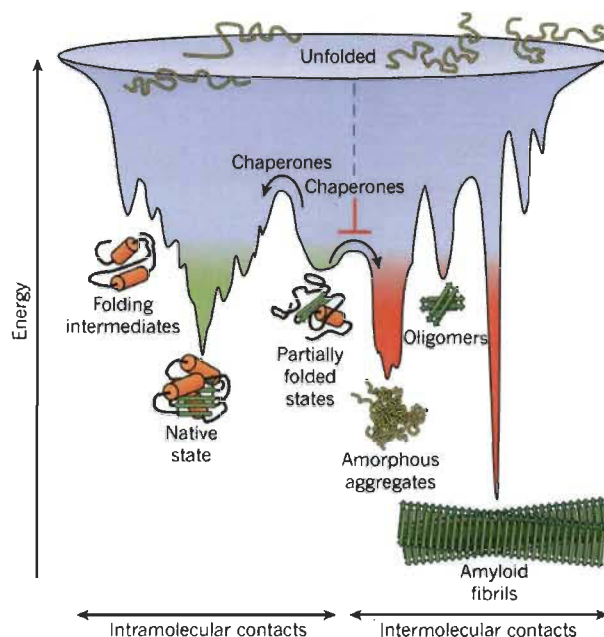


Figure 1.22 Schématisation théorique des différents paysages énergétiques associés aux différents états des protéines.

États énergétiques à la base des interactions intramoléculaires impliquées dans le repliement et dans la stabilité des protéines (à gauche) et des interactions intermoléculaires impliquées dans la formation d'agrégats amorphes ou d'oligomères ordonnés comme les fibrils ou les cristaux (Harti *et al.*, 2011).

1.7.2 Modélisation de la structure des protéines et annotation génomique

Le volume croissant de nouvelles séquences déposées annuellement dans les bases de données disponibles sur le web dépasse largement le nombre de protéines pour lesquelles la structure est connue. Environ seulement 1 % du nombre de séquences

déposées dans UniProtKB ont une structure élucidée et déposée dans PDB (Ridgen 2009). Aujourd'hui, plus de 80,000 structures de protéines ont été élucidées et déposées dans PDB et il semble qu'un nombre limité de repliements soit adopté par celles-ci. Environ 1200 types de repliements sont maintenant répertoriés dans PDB. Vu le lot d'information utilisable, le problème peut donc s'inverser : on peut chercher à déterminer quel repliement parmi ces types limités de repliements est adopté par la séquence d'une nouvelle protéine donnée. L'alignement multiple de différentes séquences adoptant le même repliement peut permettre d'identifier des patrons de similitude entre ces différentes séquences. Les travaux de David Eisenberg ont démontré que la conservation du patron d'hydrophobicité entre deux protéines est un bon indicateur de leur similarité structurale (Eisenberg *et al.*, 1982, 1984, 1987). À cet égard, Bowie *et al.*, ont démontré que le patron d'hydrophobicité/hydrophilicité de différentes séquences adoptant le même repliement peut être utilisé pour prédire le repliement global d'une nouvelle protéine sur la base d'une comparaison de séquences (Bowie *et al.*, 1990). Différents algorithmes ont été développés ces dernières années pour relier avec le plus de précision possible l'espace à une dimension des séquences d'acides aminés à l'espace à trois dimensions formé par la conformation (Bowie *et al.*, 1989, 1991; Fischer et Eisenberg 1996).

Le terme « problème du repliement inversé » est utilisé pour définir différentes techniques permettant de déterminer si la séquence d'une protéine est compatible avec la structure d'un repliement donné en définissant des critères d'empaquetage ou de préférence de contact par exemple. Les techniques permettant d'obtenir de l'information sur la structure des protéines, telles la cristallographie et la résonance magnétique nucléaire, présentent certaines limitations, sont exigeantes et beaucoup trop lentes, ce qui ralentit grandement la détermination de nouvelles structures et la compréhension des mécanismes par lesquels les protéines adoptent leurs structures tridimensionnelles. La modélisation de la structure d'une nouvelle protéine sur la base de son homologie de séquence avec une protéine homologue pour laquelle la structure est connue constitue une alternative intéressante et plus rapide pour estimer la structure globale et le repliement d'une nouvelle protéine. Cependant, la fidélité du modèle obtenu par ces

différentes approches d'homologie (fidélité des rotamères et fines interactions au sein de la protéine) repose sur le haut niveau d'identité entre la protéine modélisée et la structure cristalline utilisée comme gabarit.

1.7.3 Design de protéines et prédiction *ab initio*

Les nouvelles technologies et l'évolution de la force de calcul des ordinateurs ont permis l'avènement d'une nouvelle approche de design de protéine. Cette méthode consiste à concevoir une nouvelle protéine ou à modifier un domaine de celle-ci par une approche computationnelle. Le design computationnel s'apparente à l'évolution accélérée, mais il s'effectue de façon virtuelle. Il consiste en la recherche de séquences protéiques qui adoptent la structure voulue, définie comme un minimum énergétique. En ce sens, c'est une approche inverse à la prédiction de structure. Le design de protéine peut s'appliquer pour faire évoluer différentes propriétés d'une enzyme, comme sa stabilité, sa spécificité ou sa fonction. Cette approche est très prometteuse et plusieurs protéines ont évolué avec succès dans une direction voulue par cette approche computationnelle. L'équipe de Stephen Mayo à Caltech a réalisé plusieurs avancées majeures à cet égard (Dahiyat et Mayo 1996, 1997; Gordon *et al.*, 1999; Bolon *et al.*, 2002; Marshall et Mayo 2001; Chica *et al.*, 2010; Privett *et al.*, 2011).

Lors d'un évènement initié en 1994, Moult et ses collègues fondent CASP « Critical Assessment of protein Structure Prediction ». CASP est une communauté de membres participants à une compétition à l'aveugle dans laquelle des centaines de séquences cibles pour lesquelles la structure est connue, mais pas encore publiée, sont présentées aux participants afin que ceux-ci puissent tenter de déterminer, *ab initio*, la structure de ces séquences à l'aide de leurs nouveaux algorithmes. Lors de la neuvième édition, CASP9, deux algorithmes ont retenu l'attention, HHPred (Hildebrand *et al.*, 2009) et Rosetta (Khulman *et al.*, 2003). Aujourd'hui, la conception de nouvelles protéines *de novo*, repliables est devenue un travail de routine (Simons 1999; Dantas 2003; Jiang *et al.*, 2008; Kulik 2012). L'équipe de David Baker a réalisé plusieurs avancées dans ce domaine ces dernières années. Plusieurs nouvelles protéines

artificielles de conception *de novo*, ont été produites, cristallisées et leurs structures déterminées avec succès. La protéine Top 7 est un bon exemple et constitue la première protéine *de novo* adoptant un nouveau repliement, n'ayant pas encore été identifié dans la nature (Kuhlman et Baker 2003). Le succès de l'algorithme Rosetta a amené Baker à développer « Foldit », un jeu vidéo en ligne par lequel les joueurs sont invités à participer au développement d'algorithmes en essayant de replier une séquence protéique (séquence pour laquelle la structure est élucidée, mais inaccessible aux joueurs). Les joueurs ayant réussi à obtenir le plus de « hits » ou le meilleur « score » en repliant la séquence, obtenant la meilleure structure finale dans le plus court laps de temps (par le plus court chemin) font partie d'un classement. Les informations des meilleurs chemins choisis par ces joueurs sont recueillies et utilisées afin d'étudier les similitudes et de concevoir de nouveaux algorithmes (Cooper *et al.*, 2010; Khatib *et al.*, 2011). Le concept d'entonnoirs et les conjectures de Baldwin ont permis de développer différents algorithmes de prédiction de structures. À l'heure actuelle, deux principales approches sont considérées pour aborder le problème du repliement : l'approche purement codique qui cherche à savoir s'il existe un code pour le repliement (message du repliement encodé dans la séquence) d'une protéine et l'approche par minimisation d'énergie, utilisant les différents concepts de paysage énergétique et de champs de forces afin d'améliorer le succès de prédiction des algorithmes.

Afin de développer un outil permettant de prédire la structure native d'une protéine, avec plus de précision que les simulations de Monte-Carlo ou les différentes méthodes de dynamique moléculaire et de simulation actuelles, différentes informations issues de l'expérimentation devront s'accumuler afin d'implémenter de nouveaux concepts aux algorithmes actuels ou de développer de nouvelles approches (Dill 2007). La biologie est différente de la physique sur ce point : il est difficile de développer des lois sur une base purement théorique. Est-ce que seule l'expérimentation peut permettre de résoudre les problèmes complexes en biologie? Rappelons-nous que les différents modèles théoriques développés pour l'élucidation du code génétique ont tous échoués, ce dernier étant dégénéré, seule l'expérience a pu permettre de résoudre le problème.

1.7.4 Espace de séquence et explosion combinatoire

Si un code existe pour prédire le repliement d'une protéine, on peut imaginer une expérience de pensée, laquelle consisterait à substituer chacun des acides aminés d'une protéine par chacun des 19 autres acides aminés entrants dans l'alphabet des protéines, et ce de façon combinatoire et de mesurer l'impact de chaque combinaison sur le repliement, la stabilité ainsi que sur d'autres propriétés. Le problème est que cette expérience est absurde puisque physiquement impossible, l'espace de séquence des protéines étant beaucoup trop vaste.

Par exemple, afin d'explorer l'espace de séquence d'une protéine de longueur modeste, soit 200 résidus, un total de 20^{200} ou 1.6×10^{260} combinaisons existent. Le nombre d'atomes dans l'univers observable étant estimé à environ 10^{80} , la coexistence de 1.6×10^{260} séquences de 200 acides aminés nécessiteraient plus de 5.2×10^{263} atomes, excédant par 5.2×10^{183} fois la quantité d'atomes contenus dans l'univers observable. Le temps nécessaire à produire ce nombre de séquences utilisant un nombre limité d'atomes reste tout aussi absurde. On peut penser à analyser ces séquences en simulation informatique, mais encore ici, prenant en compte la capacité de stockage de 2.4×10^{21} bits d'information et la capacité de calcul de 6.4×10^{18} informations par secondes et ce en utilisant tous les ordinateurs existant sur la planète à l'heure actuelle, il serait impossible de simplement stocker cette information (Lloyd 2000; Hilbert et López, 2011). Imaginant une capacité de stockage suffisante, l'analyse de ces séquences avec la capacité de calcul actuelle prendrait 7.97×10^{233} ans, soit 6.14×10^{223} fois l'âge de l'univers. Cette explosion combinatoire démontre bien la nécessité d'utiliser d'autres approches pour comprendre comment la séquence des protéines détermine leur repliement. Comment et dans quelle mesure la nature a-t-elle exploré l'ensemble ou une partie de ces combinaisons de séquences durant l'évolution? La clé pour comprendre les déterminants évolutifs des différents types de repliements et de leur stabilité doit se trouver au niveau de régions plutôt discrètes (points d'ancrage) localisées au niveau de la séquence des protéines. La détermination d'un repliement par la distribution particulière de ces points d'ancrage peut réduire de façon significative l'exploration de l'espace de séquence et expliquer qu'un nombre limité de repliements peut exister.

Quoi qu'il en soit, il semble que ce ne soit pas nécessairement l'ensemble de la séquence (la totalité des résidus) qui soit essentiel pour déterminer le repliement global d'une protéine, puisque plusieurs protéines présentant une très faible identité de séquence (voire $> 25\%$) peuvent adopter le même repliement. Les éléments de structures secondaires peuvent être formés par des séquences très différentes (certaines séquences de résidus ont une propension à former des hélices et d'autres, des brins bêta par exemple). Par ailleurs, quelques mutations à des endroits clés de la séquence d'une protéine donnée peuvent entraîner d'importants changements structuraux ou même empêcher son repliement. Certains résidus ou motifs dans la séquence doivent favoriser certaines interactions essentielles à l'empaquetage de la chaîne polypeptidique d'un repliement particulier, ceux-ci pouvant servir de points d'ancrage pour initier le processus rapide du repliement de la chaîne polypeptidique suivant un chemin d'énergie préférentiel. Les patrons d'hydrophobicité et d'hydrophilicité sont des bons indicateurs d'un repliement particulier. Comme les effets hydrophobes, les effets de répulsion électrostatique peuvent jouer un rôle crucial dans le repliement et la formation de la structure tertiaire d'une protéine, la mutation d'un ou plusieurs de ces résidus peut s'avérer critique pour l'atteinte de la structure native.

1.8 Évolution dirigée et étude de la thermostabilité par mutagenèses rationnelles

Le repliement α/β hydrolases constitue un repliement protéique de choix pour l'étude de la stabilité, entre autres pour l'étude de la thermostabilité. Une large variété d'enzymes adopte ce repliement et la thermostabilité est une propriété qui varie beaucoup au sein des enzymes lipolytiques. Afin de comprendre ou d'améliorer une propriété particulière d'une enzyme donnée comme sa thermostabilité, différentes approches peuvent être utilisées.

1. L'évolution accélérée est une des méthodes les plus efficaces pour améliorer et adapter les propriétés d'une enzyme aux conditions ciblées pour une application biotechnologique spécifique. En ce sens, les propriétés sont adaptées spécifiquement aux conditions ciblées et la même propriété peut ne pas être

adaptée de façon plus générale, par exemple à des conditions légèrement différentes à celles ciblées. Par évolution accélérée, on obtient les propriétés pour lesquelles la méthode de criblage ou de sélection a été développée et le mot d'ordre est « You get what you screen for ». La connaissance de la structure tridimensionnelle n'est pas nécessaire pour l'évolution d'une enzyme par cette méthode qui a l'avantage de s'opérer directement en criblant les versions du gène dont le produit présente les propriétés enzymatiques recherchées. Puisque la sélection est effectuée par la détection réelle (chimique) de la propriété recherchée, au niveau de la protéine exprimée physiquement, l'évolution accélérée est une des méthodes les plus puissantes et ayant eu le plus de succès pour l'ingénierie efficace d'enzymes et d'applications à l'échelle industrielle. La résolution de variants issus d'évolution accélérée peut aussi permettre d'identifier des éléments structuraux responsables de l'amélioration de la propriété désirée, telle que la thermostabilité, ces éléments étant souvent difficile à prédire par approche rationnelle. L'évolution accélérée nous a appris au fil du temps que de subtils changements au niveau de la structure, souvent le résultat de simples mutations ponctuelles, peuvent entraîner une modification exceptionnelle de ses propriétés, comme sa thermostabilité.

Les méthodes de design ont évolué et se sont développées obtenant un certain succès pour prévoir les propriétés d'une conception protéique particulière. Il est possible de prédire *in silico* une série de mutations stabilisantes sans compromettre l'activité de l'enzyme. Cette approche comporte tout de même certains risques. Une trop grande stabilité conformationnelle peut entraîner une diminution, voire une perte d'activité aux températures d'applications, due à une trop grande rigidité structurale. La thermostabilité d'une enzyme doit être adaptée à son activité. L'évolution accélérée tient compte de cet équilibre délicat et une adaptation des propriétés d'une enzyme, par exemple un accroissement de sa thermostabilité, par plusieurs rondes d'évolution accélérée a l'avantage de générer la version que l'on cherche, car c'est le hasard et la sélection du gène au niveau de la protéine exprimée physiquement qui opère. L'évolution dirigée est donc une méthode extrêmement puissante.

2. L'évolution par approche rationnelle consiste par ailleurs à faire évoluer une enzyme dans une direction voulue sur une base rationnelle. La connaissance de la structure tridimensionnelle de la protéine d'intérêt est requise par cette approche puisqu'il s'agit ici d'identifier des éléments au sein de la structure par différentes approches comparatives. L'approche rationnelle dite classique consiste à déterminer des mutations potentielles sur la base d'une comparaison avec une structure homologue en vue de conférer les propriétés recherchées par cette comparaison ou encore d'introduire des mutations issues d'un criblage computationnel. L'introduction d'une mutation rationnelle sur la base d'une comparaison structurale, dans le but de conférer la propriété recherchée peut s'avérer fructueuse ou encore catastrophique (pour la structure et/ou la fonction de la protéine) vu le niveau de complexité et la difficulté de prédiction des interactions complètes associées à cette mutation lors du repliement et/ou dans la structure finale.
3. L'évolution naturelle des enzymes est sans doute le meilleur guide pour la compréhension des mécanismes par lesquels la nature opère pour faire évoluer différentes propriétés des enzymes. Plusieurs séquences différentes peuvent adopter un repliement très similaire, mais leur stabilité en solution peut aussi largement varier en fonction des fines interactions au sein de ce même repliement (Sauer *et al.*, 1996). Par exemple, la comparaison de la structure d'une enzyme mésostable avec celle d'un homologue thermostable peut permettre d'identifier des éléments structuraux potentiellement stabilisants chez la version thermostable, lesquelles peuvent être introduits dans la version mésostable dans le but d'accroître la stabilité de cette dernière. L'analyse de la structure d'un groupe d'enzymes présentant une propriété similaire peut permettre d'identifier des éléments structuraux conservés au sein de ce groupe et absents des autres membres du repliement ne présentant pas cette même propriété. La mutation rationnelle (dirigée) de ces éléments et l'analyse de leurs effets sur la propriété étudiée, par exemple la thermostabilité d'un groupe d'enzymes peut permettre de déterminer le rôle que jouent ces éléments dans la thermostabilité de ce groupe d'enzymes

(issues de l'évolution naturelle). Cette approche destructive s'apparente à l'approche par design rationnel, mais il s'agit ici d'annihiler l'élément structural identifié sur une base rationnelle et susceptible de jouer un rôle dans la propriété étudiée afin de déterminer si ce dernier joue effectivement un rôle important dans cette propriété.

De ce fait, l'identification d'éléments contrôlant une propriété particulière d'un groupe d'enzymes naturelles et la conservation de ceux-ci au sein de ce groupe peut aussi constituer des éléments de séquences utiles pour l'annotation génomique et pour prédire les propriétés de nouvelles enzymes appartenant à ce groupe, par exemple le pH et la température optimale pour l'activité, la thermostabilité, etc. L'identification d'éléments structuraux responsables de conférer une propriété particulière à un groupe d'enzymes peut permettre de compléter les éléments perdus ou non accessibles par les autres approches. Ces éléments peuvent servir à la construction de « Smart libraries » qui peuvent être davantage susceptibles de générer des résultats positifs pour l'ingénierie d'une propriété particulière d'une protéine par évolution accélérée. Cette approche d'ingénierie dite semi-rationnelle peut combiner l'information issue de l'identification de résidus conservés associés à une propriété ou l'identification de « Hot spots » par différentes approches computationnelles comme point de départ pour l'évolution accélérée (Chica *et al.*, 2005). Des études structurales en relation avec les propriétés des enzymes doivent s'accumuler afin d'établir des généralités qui permettront une meilleure compréhension de la relation séquence-structure-fonction-propriétés des protéines. Il semble que ce soit un nombre limité des résidus de la séquence d'une protéine qui détermine son repliement et sa stabilité. L'identification de résidus au niveau de la séquence d'un groupe de protéines permettant de prévoir leur comportement en solution, voire leur stabilité, constitue un outil majeur pour l'identification de nouveaux catalyseurs naturels dans les bases de données, pour des applications spécifiques dans des conditions cibles. Le problème de la stabilité est donc un aspect qui doit être abordé en parallèle à celui du repliement.

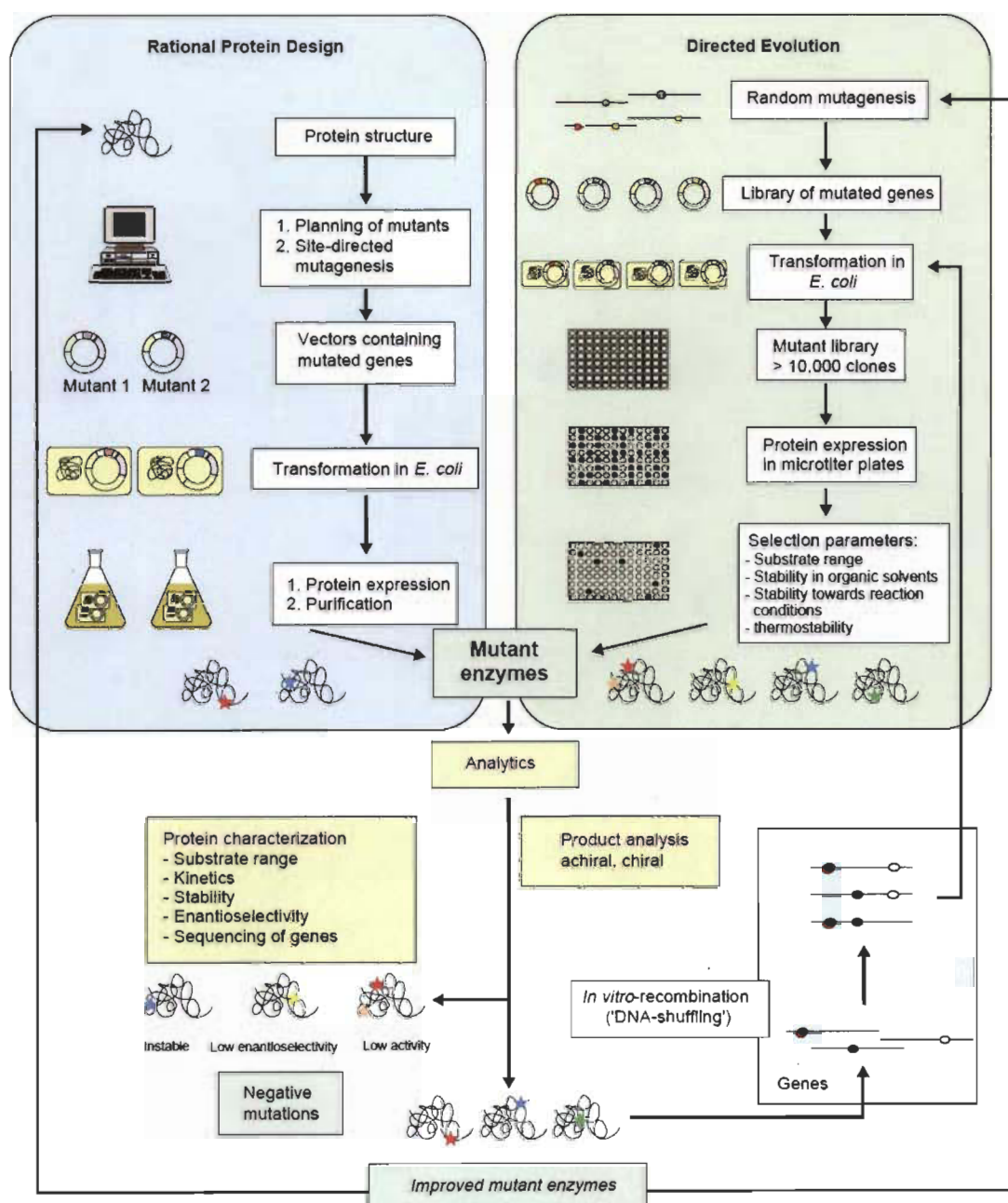


Figure 1.23 Évolution par approche rationnelle et par évolution accélérée. Principales étapes utilisées pour l'amélioration des propriétés d'une enzyme par évolution par approche rationnelle (à gauche) et par les approches aléatoires d'évolution accélérée (à droite). Figure tirée de Bornscheuer et Pohl 2001.

1.9 Objectifs

1.9.1 Objectif général

L'objectif général de cette thèse est d'identifier une nouvelle enzyme lipolytique thermostable, d'étudier certains éléments structuraux responsables de sa thermostabilité et de comprendre s'il est possible de moduler sa stabilité en fonction du pH.

1.9.2 Objectifs spécifiques

L'objectif spécifique est d'isoler et de produire une nouvelle enzyme lipolytique thermostable, de déterminer sa structure afin d'étudier et de déterminer le rôle de certains éléments contrôlant sa thermostabilité et d'utiliser ces éléments pour moduler sa réponse au pH afin d'améliorer sa thermostabilité en milieu alcalin pour des applications biotechnologiques spécifiques. On vise entre autres le traitement des extractibles dans la fabrication des pâtes et papiers, la formulation de biodétergents ou la synthèse organique en chimie fine, soit des applications biotechnologiques opérant à haute température et en milieu alcalin.

Étant donné les importantes applications industrielles que trouvent les enzymes lipolytiques, une collaboration de recherche a été entreprise dans le but d'améliorer la thermostabilité et la stabilité en milieu organique d'une nouvelle enzyme lipolytique pour la production efficace de biodiesel. L'objectif de ce projet en collaboration, était d'utiliser l'évolution accélérée pour améliorer la thermostabilité d'une enzyme lipolytique psychrostable pour la production efficace de biodiesel et de résoudre la structure cristalline des variants obtenus afin d'élucider certains éléments structuraux responsables de cette nouvelle stabilité. Ces résultats sont présentés en annexe A.

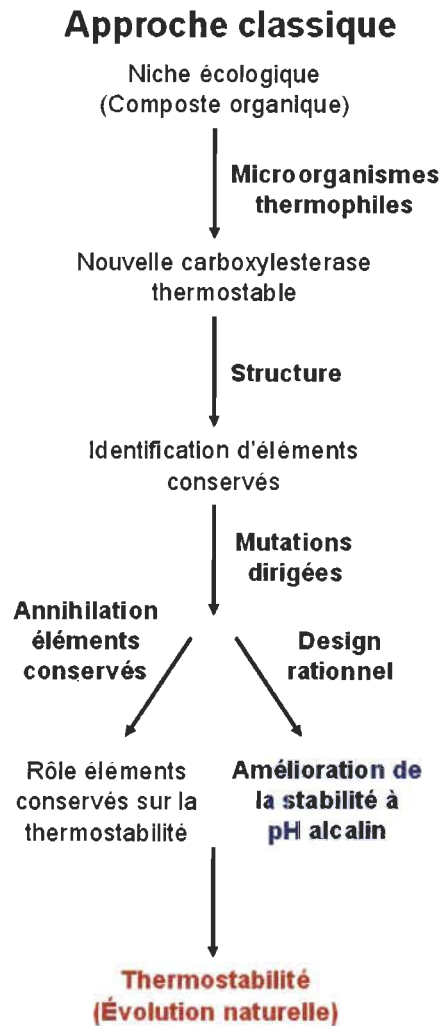


Figure 1.24 Schéma hypothétique pour l'identification de résidus contrôlant la thermostabilité d'une nouvelle enzyme lipolytique thermostable à potentiel d'application industrielle.

Une approche rationnelle sera utilisée pour identifier des éléments essentiels à la thermostabilité d'une nouvelle carboxylesterase (éléments issus d'une évolution naturelle). Le rôle potentiel que joue ces éléments sera étudié par mutagenèse dirigée et ces éléments seront aussi utilisés pour moduler la réponse au pH.

Évolution accélérée

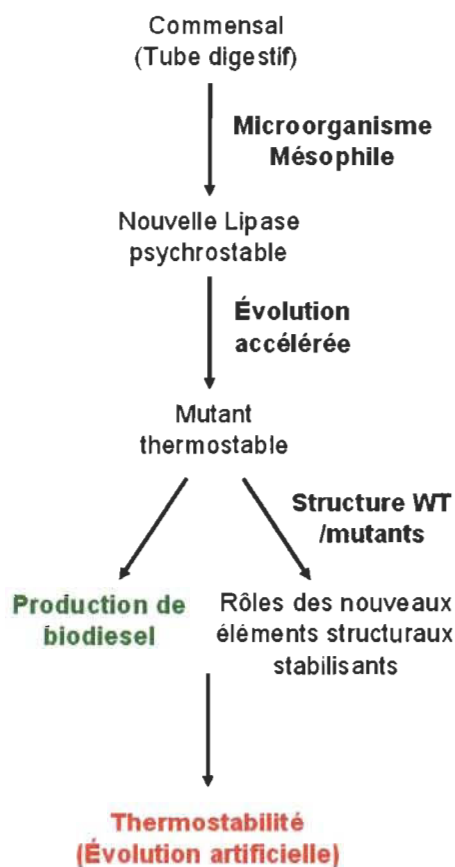


Figure 1.25 Schéma hypothétique pour l'identification de résidus contrôlant la thermostabilité de deux nouvelles enzymes lipolytiques thermostables à potentiel d'application industrielle.

Une approche aléatoire (annexe) sera utilisée pour évoluer la thermostabilité et la résistance au méthanol d'une nouvelle lipase pour la production efficace de biodiesel. L'élucidation des structures cristallines permettra d'identifier les éléments issus de cette évolution accélérée.

CHAPITRE II

IDENTIFICATION OF THERMOPHILIC BACTERIAL STRAINS PRODUCING THERMOTOLERANT HYDROLYTIC ENZYMES FROM MANURE COMPOST

Le contenu de ce chapitre est présenté sous forme de manuscrit qui a fait l'objet d'une publication dans le journal scientifique « Indian Journal of Microbiology ».

David M. Charbonneau, Fatma Meddeb-Mouelhi, Maurice Boissinot, Marc Sirois and Marc Beauregard. 2012. Identification of thermophilic bacterial strains producing thermotolerant hydrolytic enzymes in manure compost. *Indian J.* 52, 41-47

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3298584/>

2.1 Contribution des auteurs

L'ensemble des expériences à la base de cet article a été conçu par David Charbonneau qui a aussi écrit le manuscrit. Ces recherches et l'ensemble des expériences ont été menées en étroite collaboration avec Fatma Meddeb et Marc Beauregard qui ont participé à l'écriture du manuscrit ainsi qu'avec le professeur Maurice Boissinot de l'Université Laval pour ses conseils en analyses phylogénétiques et avec le professeur Marc Sirois pour son expertise en microbiologie.

2.2 Résumé de l'article

Dix souches bactériennes thermophiles ont été isolées à partir d'un échantillon de compost de fumier. Une analyse phylogénétique basée sur la séquence du gène d'ARNr 16S ainsi qu'une caractérisation phénotypique et biochimique a permis d'identifier quatre espèces appartenant à quatre genres bactériens différents : *Geobacillus*

thermodenitrificans, *Ureibacillus suwonensis*, *Bacillus smithii* et *Aneurinibacillus thermoaerophilus*. Les profils de polymorphisme de restriction (PCR-RFLP) de la séquence de l'opéron ribosomal ARNr 16S-ITS-23S a permis d'identifier deux sous-groupes dans les souches de *Geobacillus*. Les souches ont été soumises à un criblage d'activité hydrolytiques thermorésistante (60-65 °C) afin d'évaluer leur profil d'activité lipolytique (estérase et lipase), mais aussi d'autres activités hydrolytiques à potentiel d'application industrielle. Les souches de *Geobacillus thermodenitrificans*, *Bacillus smithii* et de *Aneurinibacillus thermoaerophilus* ont montré différentes activités lipolytiques thermorésistantes. Les souches de *Geobacillus* ont aussi montré de l'activité protéase, α -amylase et xylanase. Les résultats suggèrent que ces différentes espèces constituent des sources intéressantes pour l'isolation de nouvelles enzymes lipolytiques thermostable à potentiel d'application industrielle.

2.3 Premier article scientifique

Identification of thermophilic bacterial strains producing thermotolerant hydrolytic enzymes from manure compost

David M. Charbonneau[†], Fatma Meddeb-Mouelhi^{†,§}, Maurice Boissinot[‡], Marc Sirois[†]
and Marc Beauregard^{†,*}

[†] Département de Chimie-Biologie, Université du Québec à Trois-Rivières, 3351 Boul. Des Forges, C.P. 500, Trois-Rivières (Québec) G9A 5H7, Canada

[§] Buckman Laboratories of Canada, 351, Joseph-Carrier, Vaudreuil-Dorion (Québec) J7V 5V5, Canada

[‡] Centre de Recherche en Infectiologie de l'Université Laval, Centre Hospitalier Universitaire de Québec, Pavillon CHUL, 2705 Boul. Laurier, RC-709, Ste-Foy (Québec) G1V 4G2, Canada

* Corresponding author, Marc Beauregard: e-mail: marc.beauregard@uqtr.ca

Received: 16 August 2010 / Accepted: 13 January 2011 / Published online: 3 February 2011 © Association of Microbiologists of India 2011

Published in: Indian Journal of Microbiology

Indian J Microbiol 2012. 52(1):41–47 DOI 10.1007/s12088-011-0156-8

Article available online at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3298584/>

Abstract

Ten thermophilic bacterial strains were isolated from manure compost. Phylogenetic analysis based on 16S rRNA genes and biochemical characterization allowed identification of four different species belonging to four genera: *Geobacillus thermodenitrificans*, *Bacillus smithii*, *Ureibacillus suwonensis* and *Aneurinibacillus thermoaerophilus*. PCR-RFLP profiles of the 16S-ITS-23S rRNA region allowed us to distinguish two subgroups among the *G. thermodenitrificans* isolates. Isolates were screened for thermotolerant hydrolytic activities (60–65°C). Thermotolerant lipolytic activities were detected for *G. thermodenitrificans*, *A. thermoaerophilus* and *B. smithii*. Thermotolerant protease, α -amylase and xylanase activities were also observed in the *G. thermodenitrificans* group. The species represent a source of potential novel thermostable enzymes for industrial applications.

Introduction

Industrial enzymology is an important sector of biotechnology, characterised by a substantial growth over the last decade [1]. Hydrolytic enzymes have many industrial applications in the detergent sector, pulp and paper, and/or biofuel industry [2, 3]. An important source of thermostable enzymes can be found in the bacterial kingdom, where thermophilic prokaryotes are distributed among various groups in the phylogenetic tree of Bacterial and Archaeal domains [4]. Compost is a self-heating aerobic medium that results from microbial degradation of organic matter [5]. When the temperature reaches 50–55°C, the microbial thermophilic community becomes dominated by members of the genus *Bacillus* [6]. Investigation of 16S and 23S rRNA gene sequences has revealed a high heterogeneity in the genus *Bacillus* [7] and following extensive phylogenetic analysis, the genus has undergone major taxonomic revision leading to the establishment of new genera [8]. The genera, *Geobacillus* [9], *Ureibacillus* [10], and *Aneurinibacillus* [11] are outcome of this classification. Here, we report on the identification and characterization of a cultivable thermophilic community in manure compost. Novel strains were identified and further investigated for thermotolerant hydrolytic activities.

Materials and methods

Substrates and chemicals

All substrates and chemicals were of reagent grade and were purchased from Sigma–Aldrich (Oakville, ON, Canada). All culture media were from DIFCO Laboratories, Becton, Dickinson and Company (Sparks, MD 21152 USA). Unless specified otherwise, all media and substrates were autoclaved at 121°C for 20 min. ABTS, Cu²⁺, rhodamine B and sugars were filter-sterilized using 0.22 μ m filters and added to the medium at 40°C.

Source of microorganisms

Microorganisms were isolated from composted manures. The compost samples used for this research were produced by Fafard et Frères Ltée (St-Bonaventure, Qc, Canada).

Isolation and culture of thermophiles

Samples of compost 10% (w/v) (sample A: pH 7.2, sample B: adjusted at pH 5.0) were enriched for 6 h at 60°C, then plated onto nutrient-agar plates at pH 7.0 and 5.0, respectively and incubated again at 60°C for 16 h. Compost sample C (pH 7.2) was plated directly on nutrient-agar, without enrichment. Colonies from sample A, B and C were isolated on LB-agar plates at 60°C. Pure cultures were grown in Luria-Bertani broth, pH 7.0 (sample A), 5.0 (B) and 7.0 (C), respectively, at 60°C, with agitation (230 rpm) and preserved as suspension in 10% (v/v) glycerol at -80°C.

PCR, RFLP and sequencing

Genomic DNA was used as template for PCR. 16S rRNA genes were amplified using the following primers: 27F: 5'-AGATTGATCCTGGCTCAG-3' and 1522R: 5'-AAGGAGGTGATCCARCCGCA-3' [12]. The 16S-ITS-23S region was amplified using the following primer set: 27F: 5'-AGATTGATCCTGGCTCAG-3' and 23S R 5'-CCCGCTTAGATGCTTTCAGC-3' [13, 14]. PCR mixtures (50 µL) were prepared using Phusion Buffer, 200 µM dNTP and 1.25 units high fidelity Phusion DNA polymerase (New England Biolab). For both 16S and 16S-ITS- 23S regions, 0.4 µM of the respective primers was used. Initial denaturation was performed at 98°C for 30 s. Amplification was established in 30 cycles of 10 s at 98°C, 20 s at 55°C and 2.5 min at 72°C, with a final extension step at 72°C for 6 min. PCR products were purified using MinElute PCR purification kit (Qiagen). 16S rRNA genes were sequenced using an ABI Prism 3700 by the Biomolecular analysis platform (Université Laval). The 16S-ITS-23S regions (~5 kb) were digested with *EcoRI*, *HindIII*, *BamHI*/*EcoRI* endonucleases.

Analysis of 16S rRNA gene sequences

16S rRNA gene sequences were edited to start at the 5' CATGCA motif resulting in a final length of 1302 bp. Searches for sequence similarity was performed using BLAST [15] using blastn accessed via the National Center for Biotechnology Information (NCBI) website. Additional 16S rRNA gene sequences were retrieved from GenBank database [16]. Multiple sequence alignments were performed with MEGA 4.1 [17] using CLUSTAL W [18] and a phylogenetic tree was inferred using the neighbour-joining method [19]. The evolutionary distances were computed using the maximum composite likelihood model, expressed as the number of base substitution per site. The stability of relationships was assessed by a bootstrap analysis based on 1,000 resampling of the neighbour-joining dataset [20].

Bacterial characterization

Gram staining, oxidase, catalase, sugar metabolism profiles and nitrate utilization were determined for each isolate. For carbohydrate and nitrate utilization, isolates were cultured at 60°C using standard methods [21]. Durham fermentation tubes were used for detection of nitrogen gas resulting from denitrification.

Thermotolerant hydrolytic activity detection

Enzymatic activity was detected on plates consisting of: Minimal medium (Mm): K_2HPO_4 , 2.5 g/L; KH_2PO_4 , 2.5 g/L; $(NH_4)2HPO_4$, 1.0 g/L; $MgSO_4 \cdot 7H_2O$, 0.2 g/L; $FeSO_4 \cdot 7H_2O$, 0.01 g/L; $MnSO_4 \cdot 7H_2O$, 0.0007 g/L; Casamino acids, 25% (w/v); and agar, 1.5% (w/v) supplemented respectively with the following substrates: glyceryl tributyrates 1% (v/v) or olive oil 1% (v/v) containing rhodamine B 0.001% (w/v) for detection of lipolytic activity [22], starch 1% (w/v) (amylase activity) [23], carboxymethylcellulose 1% (w/v) (cellulase and xylanase activity) [24], apple pectin 1% (w/v) (pectinase activity) [25], casein 1% (w/v) (protease activity) [26] and 2,20-azino-bis(3-ethylbenzthiazoline) (ABTS) 6 mM and Cu^{2+} 300 μ M (for laccase activity). All tests were carried out at 60°C. Each medium was buffered from pH 5–12 as indicated in

the text, using the following buffers: 10 mM acetate buffer (pH 5 and 6), 10 mM Tris-HCl (pH 7–9) and 10 mM glycine-NaOH (pH 10–12).

Results and discussion

Identification of thermophilic bacteria

16S rRNA gene divergence among the isolates ranged from 0.4 to 13.9% forming four separate clusters corresponding to four genera: *Geobacillus* (I), *Bacillus* (II), *Ureibacillus* (III) and *Aneurinibacillus* (IV). Phylogenetic relationships to previously cultivated species are shown in Fig. 1. The isolates resolved into four clusters (genera) and were related to *Geobacillus thermodenitrificans* (identities of 99.4–99.7% to strain NG80-2T), *Bacillus smithii* (99.6% to strain DSM 4216T), *Ureibacillus suwonensis* (99.2–99.3% to strain 6T19T) and *Aneurinibacillus thermoaerophilus* (99.9% to strain DSM 10154T), respectively. RFLP profiles (*Eco*RI and *Hind*III) for the 16S-ITS-23S region also clearly separated the isolates into four different clusters (Fig. 2). The 16S rDNA sequences data have been registered in the GenBank database under following accession numbers: CMB-A1 (GQ293453), CMB-A2 (GQ293454), CMB-A3 (GQ293455), CMB-A4 (GQ293459), CMB-A5 (GQ293460), CMB-A6 (GQ293461), CMB-A7 (GQ293456), CMB-A10 (GQ280379), CMB-B1 (GQ293457) and CMB-C1 (GQ293458).

Characterization of thermophiles

Phenotypic characteristics for each isolate are listed in Table 1. Isolates from cluster I (*Geobacillus thermodenitrificans*) were able to carry out denitrification. Two different morphotypes were identified in the *G. thermodenitrificans* group: CMB-A1, CMB-A2 and CMB-A7 colonies were rough and slightly golden colour while CMB-A3 and CMB-A10 were smooth and creamy white colour (Table 1). Differences in sugar metabolism were also observed. Strains CMB-A1 and CMB-A2 used arabinose but did not use lactose, while CMB-A3 and CMB-A10 used lactose but did not use

arabinose. These results are different from the strain *Bacillus thermodenitrificans* DSM 465T which metabolizes both lactose and arabinose [27]. Sugar metabolism profiles for CMB-A1, CMB-A2 and CMB-A7 were closer to that previously reported for *G. thermodenitrificans* NG80-2 [28, 29] whereas CMB-A3 and CMB-A10 differed from NG80-2 in terms of utilization of arabinose. Furthermore, RFLP analysis revealed differences in the *EcoRI/BamHI* digestion pattern of the 16S-ITS-23S region within the *G. thermodenitrificans* isolates. Specifically, CMB-A3 and CMB-A10 lacked a 1.4 kb restriction fragment found in CMB-A1 and CMB-A2 (Fig. 3). Together, these features resolve these isolates into two groups in the *G. thermodenitrificans* cluster I, each group likely representing distinct strains of this species. The isolate from cluster II identified as *Bacillus smithii* only utilized D-xylose, D-glucose and D-fructose. At variance, *Bacillus smithii* SC-J05-1 was reported to metabolize a large number of simple sugars [30]. Isolates from cluster III (CMB-A4 and CMB-A5) related to *U. suwonensis* formed colonies which spread over the surface of the plate. They did not metabolize any sugars tested in this study (Table 1), in agreement with previous results [31]. The remaining isolate from cluster IV, CMB-C1, was invasive and characterized by fractal growth. CMB-C1 did not metabolize any of the tested sugars, which is different from that reported for *A. thermoaerophilus* sp. nov., strain L420-91T = DSM 10154T which metabolizes D-glucose [32].

Thermotolerant hydrolytic activities

Each isolate was screened for thermotolerant hydrolytic activities at 60°C. The Cluster I isolates related to *G. thermodenitrificans* (CMB-A1, A2, A3, A7 and A10) exhibited high extracellular esterase activity on tributyrin from pH 5.0 to 12.0 (see Table 2). These strains also showed some activity on olive oil from pH 6 to 9.5, as well as thermotolerant extracellular protease activity from pH 8.0 to 10.0 and α -amylase from pH 6.0 to 8.0. In contrast, the isolate from cluster IV, CMB-C1 related to *Aneurinibacillus thermoaerophilus*, showed intense lipolytic activity on olive oil from pH 5–9. Isolate from cluster II CMB-B1 related to *Bacillus smithii*, also exhibited lipolytic activity on olive oil plates from pH 5.0 to 7.0 while isolates from cluster III,

CMB-A5 and A6 related to *Ureibacillus suwonensis* exhibited no extracellular thermotolerant hydrolytic activity on any substrate tested here. Lipases and esterases have been reported and characterized for different *Bacillus* species including: *G. stearothermophilus*, *G. thermoleovorans* and *G. kaustophilus*. However, no lipolytic enzymes have been reported as of yet for *G. thermodenitrificans*. Due to relatedness of CMB-A2 to strain NG80-2 (genome sequence ID: NC_009328), we identified genes that we used for designing primers. One of these genes coded for a lipase/esterase (GeneID: 4966646). We amplified this locus from CMB-A2, cloned, sequenced and functionally expressed the related gene in *E. coli*. Recently we reported on the detailed characterization of this enzyme (EstGtA2) from *G. thermodenitrificans* CMB-A2 [33]. We observed 99.1% identity with the respective enzyme sequence predicted in the genome of strain NG80-2. This result corroborates the relatedness of the isolates from cluster I to *G. thermodenitrificans*. In a recent paper, a lipase produced by *A. thermoaerophilus* strain HZ was partly characterised [34]. The pH range reported for optimal lipase activity is compatible with the one reported here. Unfortunately, considering the conciseness of the phylogenetic analysis performed for strain HZ, one cannot ascertain its relatedness to the strain CMB-C1.

Acknowledgments

This work was supported through grants from NSERC. D.M.C. acknowledges the support of a PROTEO scholarship. Special thanks to Sandra Isabel for providing guidance in phylogenetic analysis, and to Dr Martin Kalmokoff (Agriculture Canada) for editorial review of the manuscript.

References

1. Freedonia Group (2010) World enzyme business report. Freedonia Group Inc, Cleveland
2. Gupta R, Gupta N, Rath P (2004) Bacterial lipases: an overview of production, purification and biochemical properties. *Appl Microbiol Biotechnol* 64:763–781
3. Hasan F, Shah AA, Hameed A (2005) Industrial applications of microbial lipases. *Enzyme Microb Technol* 39:235–251
4. Kristjansson JK, Stetter KO (1992) Thermophilic bacteria. In: Kristjansson JK (ed) *Thermophilic bacteria*. CRC press, Boca Raton, pp 1–18
5. Bertoldi M, Vallini G, Pera A (1983) The biology of composting: a review. *Waste Manage Res* 1:157–176
Claus D, Berkeley RCW (1986) Genus *Bacillus* Cohn. In: Sharpe
6. NS, Holt JG (eds) *Bergey's manual of systematic bacteriology* vol. 2. Williams and Wilkins, Baltimore, pp 1121–1125
7. Ash C, Priest FG, Collins MD (1993) Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test proposal for the creation of a new genus *Paenibacillus*. *Antonie van Leeuwenhoek* 64:253–260
8. Fritze D (2004) Taxonomy of the genus *Bacillus* and related genera: the aerobic endospore-forming bacteria. *Phytopathology* 94:1245–1248
9. Wainø M, Tindall BJ, Schumann P, Ingvorsen K (1999) *Gracilibacillus* gen. nov., with description of *Gracilibacillus halotolerans* gen. nov., sp. nov.; transfer of *Bacillus dipsosauri* to *Gracilibacillus dipsosauri* comb. nov., and *Bacillus salexigens* to the genus *Salibacillus* gen. nov., as *Salibacillus salexigens* comb. nov. *Int J Syst Bacteriol* 49:821–831
10. Fortina MG, Pukall R, Schumann P, Mora D, Parini C, Manachini PL, Stackebrandt E (2001) *Ureibacillus* gen. nov., a new genus to accommodate *Bacillus thermosphaericus* (Anderson *et al.*, 1995), emendation of *Ureibacillus thermosphaericus* and description of *Ureibacillus terrenus* sp. nov. *Int J Syst Evol Microbiol* 51: 447–455
11. Shida O, Takagi H, Kadowaki K, Komagata K (1996) Proposal for two new genera, *Brevibacillus* gen. nov. and *Aneurinibacillus* gen. nov. *Int J Syst Bacteriol* 46:939–946

12. Zhang YC, Ronimus RS, Turner N, Zhang Y, Morgan W (2002) Enumeration of thermophilic *Bacillus* species in compost and identification with a random amplification polymorphic DNA (RAPD) protocol. *Syst Appl Microbiol* 25:618-626
13. Abd-El-Haleem D, Layton CA, Sayler GS (2002) Long PCR amplified rDNA for PCR-RFLP- and Rep-PCR based approaches to recognize closely related microbial species. *J Microbiol Methods* 49:315-319
14. Terefework GNZ, Suomalainen S, Paulin L, Lindstrom K (1998) Phylogeny of *Rhizobium galegae* with respect to other rhizobia and agrobacteria. *Int J Syst Bacteriol* 48:349-356
15. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TM (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res* 36:W5-W9
16. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (2007) GenBank. *Nucl Acids Res* 36:D25-D30
17. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596-1599
18. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice. *Nucl Acids Res* 22:4673-4680
19. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425
20. Swofford DL, Sullivan J (2002) Phylogeny inference based on parsimony and other methods using PAUP. In: Salemi M, Vandamme AM (eds) *The phylogenetic handbook. A practical approach to DNA and protein phylogeny*. Cambridge University press, Cambridge, pp 160-202
21. Holt JG, Krieg NR, Sneath PHA, Staley JT, Williams ST (1994) *Bergey's manual of determinative bacteriology*, 9th edn. Williams & Wilkins, Baltimore, p 787
22. Kouker G, Jaeger K-E (1987) Specific and sensitive plate assay for bacterial lipases. *Appl Environ Microbiol* 53:211-213
23. Fuwa H (1954) A new method for microdetermination of amylase activity by the use of amylase as the substrate. *J Biochem* 41: 583-603

24. Teather RM, Wood PJ (1982) Use of congo red-polysaccharide interaction in enumeration and characterisation of cellulotic bacteria from bovine rumen. *Appl Environ Microbiol* 43:777–780
25. McKay AM (1988) A plate assay method for the detection of fungal polygalacturonase secretion. *FEMS Letters* 56:355–358
Cheeseman GC (1963) Action of rennet and other proteolytic enzymes on casein in casein-agar gels. *J Dairy Res* 30:17
26. Manachini PL, Mora D, Nicastro G, Parini C, Stackebrandt E, Purkall R, Fortina MG (2000) *Bacillus thermodenitrificans* sp. nov., nom. Rev *Int J Syst Evol Microbiol* 50:1331–1337
27. Wang L, Tang Y, Wang S, Liu L-R, Liu M-Z, Zhang Y, Liang L-F, Feng L (2006) Isolation and characterization of a novel thermophilic bacterial strain degrading long chain an-alkanes. *Extremophiles* 10:347–356
28. Feng L, Wang W, Cheng J, Ren Y, Zhao G, Gao C, Tang Y, Liu X, Han W, Peng X, Liu R, Wang L (2007) Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80–2 isolated from a deep-subsurface oil reservoir. *Proc Natl Acad Sci* 104:5602–5607
29. Takashima Y, Yamaga Y, Mitsuda S (1998) Nitrile hydratase from a thermophilic *Bacillus smithii*. *J Ind Microbiol Biotechnol* 20:220–226
30. Kim BY, Lee SY, Weon HY, Kwon SW, Go SJ, Park YK, Schuman P, Fritze D (2006) *Ureibacillus suwonensis* sp. nov., isolated from cotton waste composts. *Int J Syst Evol Microbiol* 56:663–666
31. Meir-Stauffer K, Busse H-J, Rainey FA, Burghardt J, Scheberl A, Hollaus F, Kuen B, Markristathis A, Sleytr U, Messner P (1996) Description of *Bacillus thermoaerophilus* sp. nov., to include sugar beet isolates and *Bacillus brevis* ATCC 12990. *Int J Syst Bacteriol* 46:532–541
32. Charbonneau DM, Meddeb-Mouelhi F, Beauregard M (2010) A novel thermostable carboxylesterase from *Geobacillus thermodenitrificans*: evidence for a new carboxylesterase family. *J Biochem* 148:299–308
33. Masomian M, Rhaman RNZRA, Salleh AB, Basri M (2010). A unique thermostable and organic solvent tolerant lipase from newly isolated *Aneurinibacillus thermoaerophilus* strain HZ: physical factor studies. *World J Microbiol Biotechnol* 26: 1693–1701

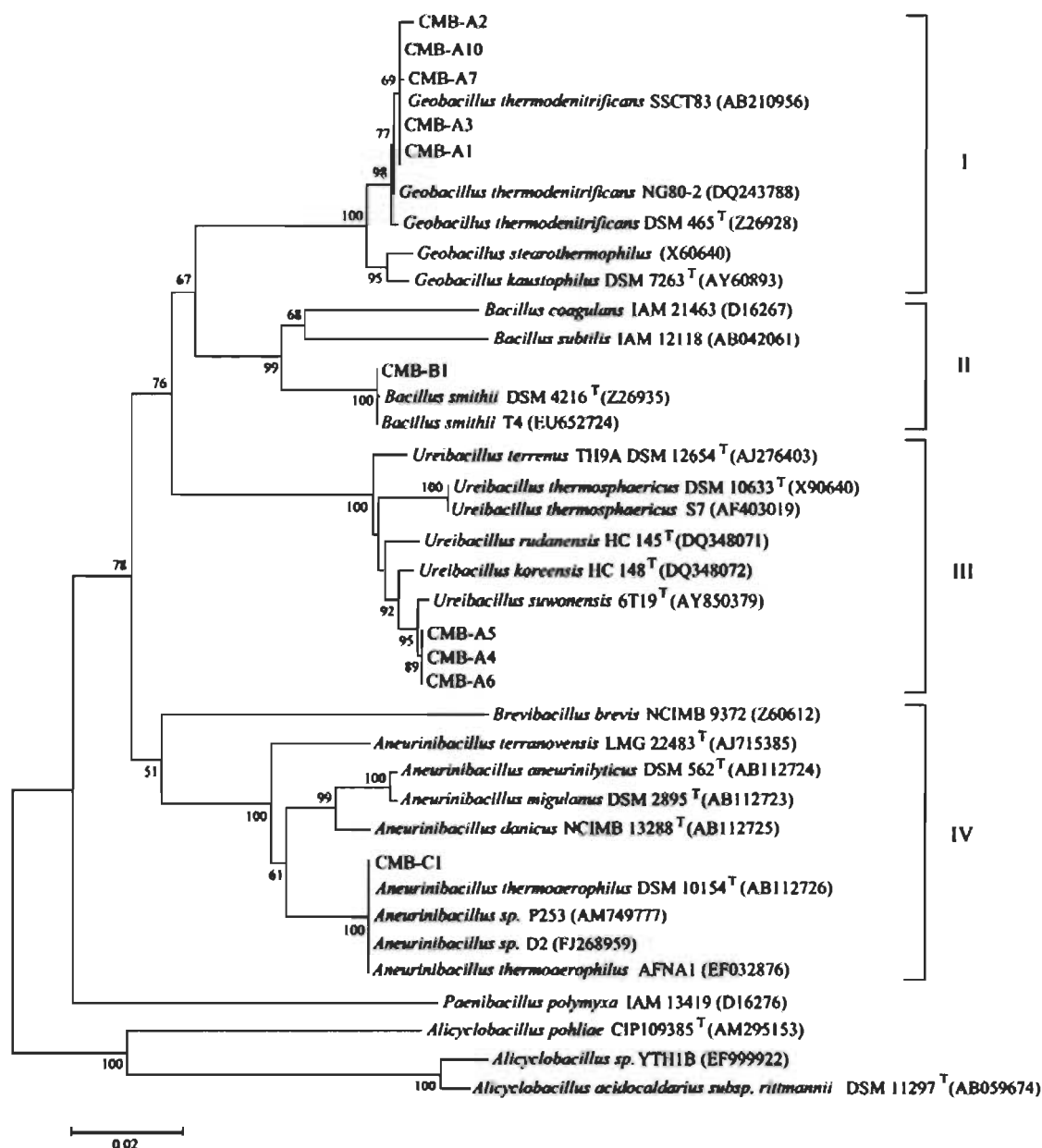


Figure 1 Neighbour-joining inferred tree based on 16S rRNA gene sequences showing the evolutionary relationship of isolates within previously characterized species. The genus *Alicyclobacillus* was used as the 'outgroup'. Type strains are indicated by a T in superscript. GenBank accession numbers of 16S rDNA sequences are shown in parentheses. Bar length represent the number of base substitution per site. Bootstrap values expressed as percentage of 1,000 resamplings of the neighbour-joining dataset are showed at nodes. Bootstrap values less than 50% are not shown.

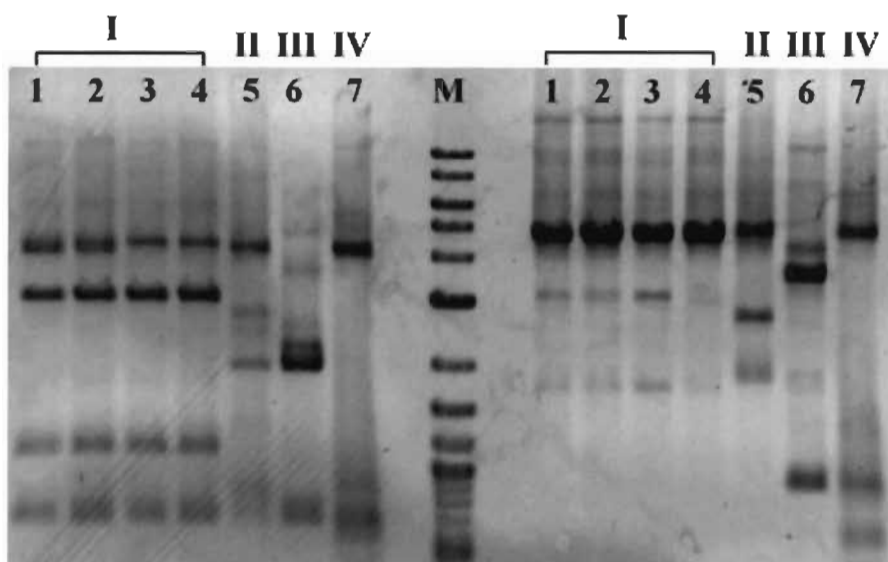


Figure 2 RFLP profiles of the 16S-ITS-23S region. Lane 1: CMB-A1; lane 2 (CMB-A2), lane 3 (CMB-A3), lane 4 (CMB-A10), lane 5 (CMB-A4), lane 6 (CMB-B1) and lane 7 (CMB-C1). M: molecular weight marker (NEB 2 log DNA ladder, 0.1–10 kbp). Roman numerals (I, II, III and IV) indicate respective clusters. Left and right side of the marker shows *EcoRI* and *HindIII* restriction profiles, respectively.

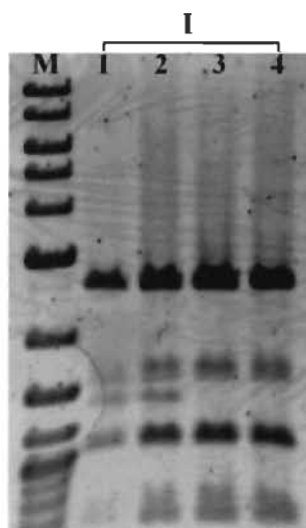


Figure 3A RFLP profiles of the 16S-ITS-23S region for *Bam*HI/*Eco*RI. Lane 1 (CMB-A1), 2 (CMB-A2), 3 (CMB-A3), 4 (CMB-A10), *M* molecular weight marker (2 log DNA ladder).



Figure 3B Microscopie optique. Colloration Gram et visualisation microscopique des cellules de *Geobacillus thermodenitrificans* CMB-A2.

Table 1 Phenotypic and biochemical characteristics of thermophilic isolates

	I	I	I	I	I	II	III	III	III	IV
Characteristic	CMB-A1	CMB-A2	CMB-A3	CMB-A7	CMB-A10	CMB-B1	CMB-A4	CMB-A5	CMB-A6	CMB-C1
Colonial Morphology	Rough	Rough	Smooth	Rough	Smooth	Glistening	n/a ¹	n/a ¹	n/a ¹	Rough
Colonic colour	Golden	Golden	Cream	Golden	Cream	White	White	White	Brown	White
Cell shape	Rods	Rods	Rods	Rods	Rods	Rods	Rods	Rods	Rods	Rods
Gram stain	+	+	+	+	+	+	—	v	v	+
KOH (3%) test	—	—	—	—	—	—	—	—	—	—
Catalase activity	+	+	+	+	+	+	+	+	+	—
Oxydase activity	—	—	—	—	—	—	+	+	+	—
Denitrification	+	+	+	+	+	—	—	—	—	—
Growth temp range (°C)	45–65	50–65	50–65	50–60	50–65	45–65	50–60	50–60	50–65	45–60
Growth pH range ^a	7–9.5	7–9.5	7–9.5	7–9.5	7–9.5	5–7.5	7–9.5	7	7–9	6–8
L-Arabinose	+	+	—	+	—	—	—	—	—	—
D-Cellobiose	+	+	+	+	+	—	—	—	—	—
Lactose	—	—	+	+	+	—	—	—	—	—
D-Raffinose	—	—	—	w	—	—	—	—	—	—
D-Xylose	+	+	+	+	+	+	—	—	—	—
α-D-Sucrose	+	+	+	+	+	—	—	—	—	—
D-Sorbitol	—	—	—	—	—	—	—	—	—	—
D-Glucose	+	+	+	+	+	+	—	—	—	—
D-Galactose	+	+	+	+	+	—	—	—	—	—
D-Fructose	+	+	+	+	+	+	—	—	—	—

^a LB liquid medium*i* invasive, + positive reaction, — negative reaction, w weakly positive reaction, v variable reaction. Roman number I, II III and IV represent respective cluster

Table 2 Thermotolerant extracellular hydrolytic activities for isolates within each respective cluster (I–IV)

	I	I	I	I	I	II	III	III	III	IV
Hydrolytic activities	CMB-A1	CMB-A2	CMB-A3	CMB-A7	CMB-A10	CMB-B1	CMB-A4	CMB-A5	CMB-A6	CMB-C1
Lipolytic	5–9	5–9	5–9	5–9	5–9	5–9	–	–	–	5–9
Esterase	5–13	5–12	5–12	5–12	5–12	– ^b	– ^b	– ^b	– ^b	–
Amylase	6–8 ^a	6–7 ^a	6–8 ^a	6–8 ^a	6–7 ^a	–	–	–	–	– ^b
Xylanase	6 ^a	7–8 ^a	6 ^a	–	6 ^a	6 ^a	–	–	6 ^a	–
Cellulase	– ^b	– ^b	– ^b	– ^b	– ^b	–	– ^b	–	7 ^a	– ^b
Pectinase	– ^b	– ^b	– ^b	– ^b	– ^b	–	–	–	–	–
Protease	8–10	8–10	8–10	8–10	8–10	–	–	–	–	– ^b
Laccase	– ^b	– ^b	– ^b	– ^b	– ^b	–	–	–	–	–

^a Activity at 50°C

^b Growth with substrate as sole carbon source

– No activity detected

Numbers show pH values where hydrolytic activity were detected at 60°C

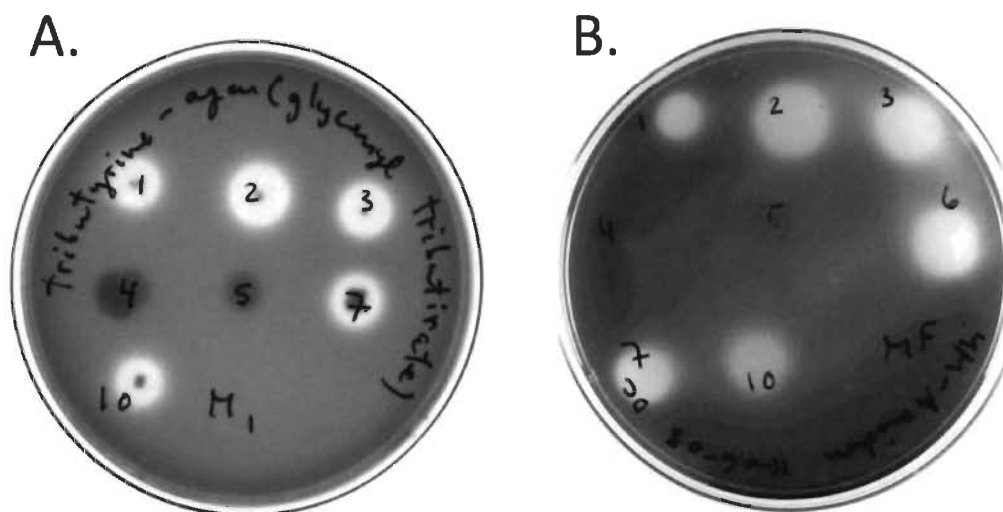


Figure 4 Detection and screening of esterase and α -amylase activity. A. The isolates were screened on emulsified agar plates containing 1% tributyrin (A) or 1% starch (B). The clear zones surrounding colonies indicate the hydrolysis of the emulsified substrate.

CHAPITRE III

A NOVEL THERMOSTABLE CARBOXYLESTERASE FROM *GEOBACILLUS THERMODENITRIFICANS*: EVIDENCE FOR A NEW CARBOXYLESTERASE FAMILY

Le contenu de ce chapitre a fait l'objet d'une publication dans le journal scientifique « Journal of Biochemistry » hébergé par la presse de l'Université d'Oxford.

David M. Charbonneau, Fatma Meddeb-Mouelhi and Marc Beauregard. 2010. A novel thermostable carboxylesterase from *Geobacillus thermodenitrificans*: Evidence for a new carboxylesterase family. *J. Biochem.* 148, 299-308

<http://jb.oxfordjournals.org/content/148/3/299.long>

3.1 Contribution des auteurs

Ces recherches et l'ensemble des expériences qui ont mené à la publication de cet article ont été conçus par David Charbonneau, Fatma Meddeb et Marc Beauregard qui ont aussi participé à l'analyse des résultats. David Charbonneau a rédigé en entier le manuscrit qui a été révisé par Fatma Meddeb et Marc Beauregard avant sa publication.

3.2 Résumé de l'article

Un nouveau gène encodant une estérase de *Geobacillus thermodenitrificans* CMB-A2 a été cloné, séquencé et exprimé fonctionnellement chez *E. coli* M15. L'analyse de la séquence révèle un cadre ouvert de lecture (ORF) de 747 pb encodant un polypeptide de 249 acides aminés, appelé EstGtA2. L'enzyme purifiée montre une activité spécifique de 2,58 U/mg sur *p*NP-caprylate (C8), à 50 °C et pH 8 (conditions optimales). La nouvelle enzyme catalyse l'hydrolyse de la tributyrine et d'une variété

d'ester de *p*NP de différentes longueurs de chaînes aliphatiques (C4-C16). Cette nouvelle enzyme montre un fort potentiel pour différentes applications biotechnologiques étant donné sa versatilité de catalyse, hydrolysant une large gamme de substrats à l'intérieur d'une vaste plage de pH entre 25 et 65 °C. La structure cristalline d'une estérase homologue de *Geobacillus stearothermophilus* Est30 (PDB no. 1TQH) a servi pour modéliser la structure de EstGtA2 en utilisant le programme ESyPred3D. L'analyse de ce modèle et la comparaison des deux structures ont permis d'identifier certains éléments de structures qui peuvent être impliquées dans le contrôle des propriétés enzymatiques de EstGtA2. L'alignement multiple de séquence, une analyse phylogénétique ainsi que l'analyse des propriétés structurales et biochimiques permettent de suggérer que EstGtA2 appartient à une nouvelle famille d'enzymes lipolytiques qui diverge des huit familles originales initialement proposées par Arpigny et Jaeger.

3.3 Deuxième article scientifique

A novel thermostable carboxylesterase from *Geobacillus thermodenitrificans*: evidence for a new carboxylesterase family

David M. Charbonneau^{1,2}, Fatma Meddeb-Mouelhi^{1,3} and Marc Beauregard^{1,2,*}

Département de chimie-biologie, Université du Québec à Trois-Rivières, 3351 Boul. Des Forges, C.P. 500 Trois-Rivières (Québec) G9A 5H7

PROTEO, Université Laval, 2705 Boul. Laurier, Ste-Foy (Québec) G1V 4G2

Buckman Laboratories of Canada, 351, Joseph-Carrier Vaudreuil-Dorion (Québec) J7V 5V5, Canada

*Corresponding author: Marc Beauregard, Département de chimie-biologie, Université du Québec à Trois-Rivières, C.P. 500, Trois-Rivières (Québec), G9A 5H7, Canada, Tel: 819 376 5011 (Ext. 3354), Fax: 819 376 5084, email: marc.beauregard@uqtr.ca

Keywords: Esterase, *Geobacillus thermodenitrificans*, pNP caprylate, salt bridges, thermal stability, tributyrin.

Received April 6, 2010; accepted June 14, 2010; published online June 29, 2010

Published in *J. Biochem*: 2010;148(3):299–308 doi:10.1093/jb/mvq064

Article available online at: <http://jb.oxfordjournals.org/content/148/3/299.long>

Abstract

A novel gene encoding an esterase from *Geobacillus thermodenitrificans* strain CMB-A2 was cloned, sequenced and functionally expressed in *Escherichia coli* M15. Sequence analysis revealed an open reading frame of 747 bp corresponding to a polypeptide of 249 amino acid residues (named EstGtA2). After purification, a specific activity of $2.58 \text{ U} \cdot \text{mg}^{-1}$ was detected using *p*NP caprylate (C8) at 50°C and pH 8.0 (optimal conditions). The enzyme catalyses the hydrolysis of triglycerides (tributyrin) and a variety of *p*-nitrophenyl esters with different fatty acyl chain length (C4-C16). The enzyme has potential for various industrial applications since it is characterized by its activity under a wide range of pH, from 2.5 to 6.5°C. Using *Geobacillus stearothermophilus* Est30 esterase structure as template, a model of EstGtA2 was built using ESyPred3D. Analysis of this structural model allowed identifying putative sequence features that control EstGtA2 enzymatic properties. Based on sequence properties, multiple sequence comparisons and phylogenetic analyses, this enzyme appears to belong to a new family of carboxylesterases.

Introduction

Lipolytic enzymes are widely distributed in nature and are abundantly found in animals, plants and microorganisms (1). Carboxylesterases (EC 3.1.1.1) are hydrolases that catalyse the hydrolysis or synthesis of ester bonds in short chain aliphatic or aromatic esters (≤ 10 carbon atoms), in contrast with triacylglycerol lipases (EC 3.1.1.3) which can be defined as carboxylesterases that catalyse the hydrolysis or synthesis of long chain acylglycerol (≥ 10 carbon atoms) (2, 3). Microbial carboxylesterases have broad substrate specificity, exhibit high regio- and stereospecificity, require no cofactor and are active in organic solvents. They have important physiological, industrial and medical applications, namely in the synthesis of stereospecific compounds including the metabolic processing of drugs and antimicrobial agents (4, 5). Carboxylesterases from thermophiles also show thermal stability at elevated temperatures and have become enzymes of special interest in biotechnological applications including: detergent formulation, food processing, cosmetics, fine chemical and pharmaceutical industry (6, 7). Crystallographic structure investigations have shown that all carboxylesterases and lipases belong to the α/β hydrolase fold family (8) and bear a catalytic triad formed by Ser-Asp/Glu-His. Microbial carboxylesterases have been classified into eight families based on their sequences and biological properties (9). Here we describe the cloning and characterization of a novel thermostable carboxylesterase from *Geobacillus thermodenitrificans* named EstGtA2. We suggest that EstGtA2 is a member of a new family of carboxylesterases which diverge from families suggested by the classification of Arpigny and Jaeger (9).

Materials and methods

Isolation of bacteria from environmental samples

Geobacillus thermodenitrificans CMB-A2 was isolated in manure compost from a local distributor (Fafard et frère, St-Bonaventure, Québec, Canada). Compost was incubated at 60°C for 16 h and used to grow microbial colonies on Nutrient-agar plates. Colonies were grown at 60°C on emulsified minimal medium supplement with tributyrin

1% (v/v). Isolated colonies were identified by 16S rRNA gene sequencing as *G. thermodenitrificans* CMB-A2 (GenBank accession number GQ293454). Pure culture from single colony was cultivated in LB medium for 16 h at 60°C with agitation (230 r.p.m.) and preserved as suspension in 10% (v/v) glycerol at -80°C.

Strains, plasmids and culture conditions

Geobacillus thermodenitrificans strain CMB-A2 was used in this study. *Escherichia coli* XL1 Blue: recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacIq ZΔM15 Tn10 (Tetr)] (Stratagen®) was used for the cloning of EstGtA2 gene. *Escherichia coli* M15[pREP4] [NaI^S, Str^S, Rif^S, Thi⁻, Lac⁻, Ara⁺, Gal⁺, Mtl⁻, F⁻, RecA⁺, Uvr⁺, Lon⁺] (QIAexpress®) was used for expression of the recombinant protein. The resulting expression system named *E. coli* M15[pREP4]/pQE31-EstGtA2 was cultured in Luria Bertani broth (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl), pH 7 containing ampicillin (100 µg/mL) and kanamycin (25 µg/mL) at 37°C and 230 rpm.

Preparation of genomic DNA

Genomic DNA was isolated from CMB-A2 using a standard method (10,11). The quality of the preparation was evaluated by spectrophotometry and by agarose gel electrophoresis.

Polymerase chain reaction

Genomic DNA was used as template for polymerase chain reaction (PCR). The 16S rRNA gene was amplified using the following primers: 27F: 5'-AGATTTGATCCTGGCTCAG-3' and 1522R: 5'-AAGGAGGTGATCCARCCGCA-3' (specific for universally conserved region of bacterial 16S rRNA gene sequences). The locus tag GTNG_1710 (Gene ID 49666446) identified in *G. thermodenitrificans* NG80-2 genome (12,13) was used to design oligonucleotides for directional cloning. The following primers were synthesized: estGt-KpnI F:

5'TAATTAGGTACCTATGAAAGAACGATATCCTGTACTT 3' and est*Gt-Hind*III R: 5' TATTAAGCTTTCAAGCATGTTTGGCGAA 3' and were used for PCR. PCR mixtures were prepared in 50 µl containing 1× Taq Buffer™, 200 µmol l⁻¹ dNTP, 1.25 units Taq DNA polymerase (New England Biolab) and 0.4 µmol l⁻¹ of the respective primers. Initial denaturation was performed at 95°C for 4 min. Amplification was established in 5 cycles of 30 s at 95°C, 30 s at 50°C and 2 min at 72°C followed by 30 cycles of 30 s at 95°C, 30 s at 58°C and 2 min at 72°C. Final extension step was achieved at 72°C for 10 min. PCR products were purified using Minelute PCR purification kit (Qiagen®) and analysed on 1% agarose gel electrophoresis.

Cloning of PCR products

The purified PCR product of the EstGtA2 ORF was digested with *Kpn*I and *Hind*III restriction endonucleases for 16 h at 37°C, purified using Minelute PCR purification kit (Qiagen®) and then ligated (4 h at 16°C) into *Hind*III/*Kpn*I-digested pQE31 expression vector (Qiagen®). The pQE31-EstGtA2 recombinant plasmid was transferred by electroporation in *E. coli* XL1 Blue (electrocompetent cells). Transformed cells were selected on LB-agar with ampicillin 100 µg/mL, containing X-Gal and IPTG. Plasmid DNA isolation was performed using QIAprep™ (Qiagen®) and the integrity of the insert was tested by colony PCR using the EstGtA2 primers set and by double digestions (*Kpn*I and *Hind*III) before the plasmid was used for sequencing of the insert.

DNA sequencing and bioinformatic analysis

Both strands of the recombinant EstGtA2 gene were sequenced using an ABI Prism 3700 by the Biomolecular analysis platform (Université Laval). Sequence edition and analysis were performed with Clone Manager Professional Suite version 7.03. Sequences for comparative study were retrieved from protein and nucleotides database on the NCBI Entrez at <http://www.ncbi.nlm.nih.gov/Entrez/>. Similarity searches were performed with BLAST 2.0 program (14) in the GenBank database (15) and with BLASTP (16). The conserved domains present in the EstGtA2 amino acid sequence were analysed using the NCBI Conserved Domain Database (CDD) (17). Multiple

sequence alignments were performed with Clone Manager Professional Suites version 7.03 using a BLOSUM62 scoring matrix. The program MEGA 4.1 (18) was used for multiple sequences alignments using CLUSTAL W (19) and for construction of a phylogenetic tree using the neighbour-joining method (20). The stability of the relationships was assessed by a bootstrap analysis based on 1,000 resamplings of the neighbour-joining (21).

Structural modelling

The EstGtA2 3D structure was predicted by homology modelling using the EsyPred3D program (22). The crystal structure of carboxylesterase Est30 from *G. stearothermophilus* (PDB code 1TQH_A) was used as template (5). Image of the resulting 3D model was generated using Pymol (DeLano Scientific, Palo Alto, CA, USA) and the quality of the model regarding geometry and stereochemistry was evaluated by Whatcheck and PROCHECK analysis (23, 24) using the Swiss-model workspace (25). The evaluation of salt bridges based on the EstGtA2 model was performed with the ESBRI interface (26).

Protein expression

The pQE31-EstGtA2 recombinant plasmid was transferred by electroporation in *E. coli* M15[pREP4] electrocompetent cells. Transformed cells were selected on LB-agar with ampicillin (100 µg/mL) and kanamycin (25 µg/mL). After selection, clones of *E. coli* M15[pREP4]/pQE31-EstGtA2 were cultured as described earlier. Expression of EstGtA2 was carried out at an OD_{600 nm} of 0.6 by addition of 1 mM IPTG for 150 min at 37°C (agitation 230 r.p.m.).

Protein purification and determination

Induced cells were centrifuged at 4°C for 30 min at 4,000 g and the pellets were stored at -80°C. Cell samples (250 µL) were taken before and after induction for SDS-PAGE analysis. The His-tagged protein was purified under native conditions using

Ni-NTA™ nickel affinity resin (Qiagen®) according to manufacturer specifications at pH 8 using imidazole for elution. In order to remove salts and imidazole, purified EstGtA2 was dialysed twice against 2 l of phosphate buffer 50 mM, pH 8.0 for 48 h at 4°C using 3,500 MWCO Slide-A-Lyzer Dialysis Cassettes® (Thermo Scientific). Concentration of protein was determined by the bicinchoninic acid method (27) using bovine serum albumin as standard (Sigma-Aldrich, St-Louis, USA). The molecular weight of the proteins was determined by SDS-PAGE and proteins were detected by Coomassie blue staining following standard procedures. A low molecular range protein standard solution (Bio-Rad Laboratories Inc) was used as molecular weight marker.

Enzyme assays

Carboxylesterase activity was determined spectrophotometrically by measuring the amount of *p*-nitrophenol (*p*NP) released after hydrolysis of different *p*NP esters. The *p*-nitrophenyl esters were prepared in DMSO. The standard reaction (0.2 mL) contained 50 µM *p*NP ester, 50 mM of appropriated buffer with 14 µg of the purified enzyme. The reaction was performed in 96-well microplates using VWR heating blocks. The production of *p*NP was monitored at 405 nm using a Thermo spectrophotometer. Specific standard curves using known amount of *p*NP were measured for each set of conditions (pHs, temperatures) in order to quantify the amount of *p*NP released during the enzymatic reactions for each condition. One unit of activity was defined as the amount of enzyme releasing 1 µmol of *p*NP per min under assay conditions.

Chain length specificity, optimal pH and temperature

Chain length specificity was determined by hydrolysis of different *p*NP esters: *p*NP butyrate (C4); *p*NP caprylate (C8); *p*NP decanoate (C10); *p*NP laurate (C12); *p*NP myristate (C14); *p*NP palmitate (C16). For temperature and pH studies, the standard reaction (0.2 mL) was carried out with *p*NP caprylate (C8) as substrate since EstGtA2 activity was maximal on this substrate. The effect of temperature was studied from 25 to 80°C. Activity was measured from pH 6 to 12 using the following buffers: sodium phosphate 50 mM was used from pH 6 to 8, Tris-HCl 50 mM was used at pH 9,

CAPS 50 mM at pH 10 and sodium dibasic (Na_2HPO_4) 50 mM at pH 11 and 12 at 50°C (optimal temperature). Enzymatic activity on glyceryl tributyrates (tributyrim) was detected on agar-plate (1.5%) containing glyceryl tributyrates (1%) by appearance of a clear zone after incubation at 50°C from pH 6 to 12.

Thermal stability

Enzyme (1.5 mL) was incubated in 50 mM phosphate buffer (pH 8) in 2 mL tubes using a Nestlab-RT111 bath. The enzyme was incubated at the specified temperatures for different period of time and then the activity was measured under standard condition at 50°C using *p*NP caprylate (C8) as substrate.

Results

Sequence of EstGtA2

A thermophilic *Bacillus* strain CMB-A2 growing in the range of 50–65°C was isolated from manure compost and identified as *G. thermodenitrificans* based on 16S rRNA gene sequence [GenBank accession no. GQ293454, 99.5% identity to *G. thermodenitrificans* NG80-2 strain (12)]. The coding sequence of a putative esterase named EstGtA2 was amplified by PCR and cloned in *E. coli* M15 using the expression vector pQE31. Translation of DNA sequence revealed a 249 amino acid sequence comprising a pentapeptide formed by Gly-Leu-Ser-Met-Gly that matched the conserved pentapeptide Gly-X-Ser-X-Gly found in most esterases and lipases. Homology comparison revealed that EstGtA2 shows 99.2% identity with the amino acid sequence of a putative esterase/lipase (YP_001125819) predicted in the genome of *G. thermodenitrificans* NG80-2 (13). This enzyme corresponds to the putative gene identified in the related genome, and used for designing PCR primers (Materials and Methods section). Among the conserved domains identified in the EstGtA2 amino acid sequence we identified pfam00561 (α/β hydrolase fold family), COG1647 (esterases/lipases), and COG2267 (lysophospholipase).

Expression and purification of EstGtA2

The recombinant EstGtA2 was overexpressed in *E. coli* M15 as a 6 × His tagged fusion protein and purified under native conditions as described in ‘Materials and Methods’ section. A new single band at the expected molecular weight of about 30 kDa was detected after purification (Fig. 1). The recombinant EstGtA2 was functionally expressed in *E. coli* M15 at a yield of 30 mg/L of culture.

Characterization of EstGtA2

The optimum pH and temperature were determined using a standard assay (using *p*NP caprylate since EstGtA2 showed highest activity on this substrate, described in Materials and Methods). The enzyme has its maximal activity at pH 8 and displayed activity over a wide range of pH values, with 62% of relative activity observed at pH 6 and 36% at pH 11 (Fig. 2). EstGtA2 showed a temperature optimum of 50°C. A maximal activity (specific activity) of 2.58 U mg⁻¹ was observed at 50°C and pH 8. The enzyme retained over 90% of its maximal activity at 55 and 60°C. However, a decrease in activity was observed at 65°C (83% left), while only 15% of relative activity was observed at 75°C (Fig. 3).

Substrate specificity was tested using *p*NP esters of different chain length at pH 8 and 50°C. As shown in Fig. 4, EstGtA2 exhibited activity toward *p*-nitrophenyl esters with a broad range of acyl chain lengths. Maximal activity was observed toward *p*NP caprylate (C8). Activity decreased when the chain length increased from C8 to C16, suggesting that EstGtA2 is an esterase. Relative activity toward other substrates tested was compared to activity on *p*NP caprylate. EstGtA2 showed moderate activity toward *p*NP decanoate (65%), *p*NP laurate (58%), *p*NP myristate (36%), and *p*NP palmitate (12%). The substrate specificity was assayed from 25 to 60°C. Interestingly, relative activity toward long chain *p*-nitrophenyl esters (*p*NP myristate and *p*NP palmitate) was increased when the temperature increased to 50 and 60°C when compared with 25–40°C (data not shown). The enzyme showed also activity on glyceryl tributyrinate (tributyrin)

over a wide range of pH value (pH 6–12), but did not show any activity on longer acyl chain triglycerides (data not shown).

Thermal stability was investigated and the results showed that the enzyme was stable for >2 h at 60°C. The enzyme stability profile was similar from 50 to 65°C, however an important decrease in stability was observed at 70°C, where the enzyme retained 63% of its initial activity, down to 43% at 75°C and further to 25% at 80°C after 10 minutes (Fig. 5).

Sequence comparison of EstGtA2

Relatedness of EstGtA2 to known carboxylesterases and lipases was investigated with a BLAST analysis which indicated high sequence similarity for the esterase/lipase (YP_001125819) of *G. thermodenitrificans* strain NG80-2 (99.2% identity, gene used for designing the PCR primers), monoacylglycerol lipase (P82597) from *Bacillus. Sp* H-257 (89% identity) (28, 29) and acylglycerol lipase (YP_147178) from *G. kaustophilus* HTA426 (86% identity).

A multiple sequence alignment (MSA) was created for EstGtA2 with its closest homologues (99–86%): YP_001125819 from *G. thermodenitrificans* strain NG80-2; P82597 from *Bacillus. Sp* H-257; YP_147678 from *G. kaustophilus* HTA426 and its closest homologue (32% identity) for which a crystal structure was known, Est30 from *G. stearothermophilus* (PDB code 1TQH, hereafter named Est30Gs). The MSA shown in Fig. 6 also contains closest homologues to Est30Gs, namely EstA from *G. thermoleovorans*, GkCE from *G. kaustophilus* HTA426 (30) and Est30 from *G. thermodenitrificans* strain NG80-2 (another putative thermostable carboxylesterase from *G. thermodenitrificans* NG80-2 also named Est30) (13). The MSA suggests two distinct groups of carboxylesterases. The first group which includes EstGtA2 (top sequences) shares ~30% identity with the second group including Est30Gs (lower half sequences). Despite the low sequence identity between these two groups, conserved blocks are identified in all the sequences of the MSA (Fig. 6). The putative catalytic triad has been identified in EstGtA2 based on comparison with Est30Gs and GkCE

sequences for which residues of the catalytic triad have been determined (5, 30). In Est30Gs, the nucleophilic serine residue is located at position 94 in the pentapeptide (Gly-Leu-S94-Leu-Gly) which corresponds to position Ser97 in EstGtA2 (located in the corresponding pentapeptide Gly-Leu-Ser97-Met-Gly), the catalytic aspartate and histidine residues, located at position Asp193 and His223 in Est30Gs, correspond to Asp196 and His226 in EstGtA2. These residues are conserved in all the sequence analysed in the MSA (Fig. 6). Despite low sequence identity, structural similarity is suggested for EstGtA2 and Est30Gs. Secondary structures prediction for EstGtA2 suggests high conservation in comparison with those calculated from Est30Gs resolved structure. The alpha helices (α A, to α F) and the beta strands (β 2 to β 8) in Est30Gs appear to be highly conserved in EstGtA2. The three helices, (B_1' , D_1' and D_2'), involved in the formation of a cap structure in Est30Gs are also predicted in EstGtA2 sequence (Fig. 6).

Structure modelling of EstGtA2

A 3D model of EstGtA2 was generated by homology modelling using Est30Gs (access 1TQH_A) as template. The Ramachandran plot (not shown) indicated that 94.1% of the residues had dihedral angles predicted in the core and 5.4% in allowed regions. Bond lengths, bond angles and torsion angles in the model were evaluated using WHAT IF program. A RMS z-score for a normally restrained data set is expected to be around 1.0. The bond length and bond angles were found to deviate normally from standards values (RMS z-score of 0.93 and 1.265) respectively. The model shown in Fig. 7 suggests that EstGtA2 can adopt a 3D structure which is very similar to Est30Gs. It also suggests the existence of two distinct domains: A larger domain resembling the classical α/β hydrolase fold; and a smaller domain formed by three helices: a small helix (B_1') with two larger helix (D_1' and D_2'), involved in the formation of a cap structure which lies over the N-terminal end. The 3D model shows that the amino acid residues of the catalytic triad (Ser97, Asp196 and His226, positions suggested on the basis of the MSA, Fig. 6) are indeed located in close proximity and are arranged in catalytic triad-like configuration (Fig. 7). This catalytic triad is located at the interface of the two domains with Ser97 located at a sharp turn between the strand β 5 and helix α C, called

the nucleophile elbow. Asp196 and His226 are positioned on loops between β 7- α E and β 8- α F, respectively.

Phylogenetic tree of carboxylesterases

A phylogenetic tree of esterases and lipases (amino acid sequences) featuring some lipase families suggested from the classification of Arpigny and Jaeger (9) is shown in Fig. 8. Sequences used for MSA analysis (Fig. 6) formed two clusters which emerge from the canonical Arpigny and Jaeger families: the cluster including Est30Gs was named (N), and the cluster including EstGtA2 was named (N'). It should be noted that clusters N and N' include the sequences used to build the MSA and show around 30% identity between each other. Sequences from the new cluster N' show ~15% identity with the sequences of the other families (I–VIII).

Discussion

In this study we characterized a new lipolytic enzyme (EstGtA2) cloned from a *Geobacillus thermodenitricans* strain that we named CMB-A2. This strain screened from compost under high temperature (60°C) was identified on the basis of a 16S rRNA analysis. The EstGtA2 sequence, structure and properties were compared with lipases and esterases identified in the classification of Arpigny and Jaeger (9) in order to assign EstGtA2 to a family (I–VIII). The novel EstGtA2 is somewhat similar to family VI. This family includes enzymes of a molecular mass in the range of (23–26 kDa), with no activity toward long-chain triglycerides and ~40% similarity with eukaryotic lysophospholipases. Sequence analysis shows that EstGtA2 is a relatively small esterase (27.4 kDa), which is active on short chain triglycerides (tributylin) and contains a lysophospholipase domain, but did not display the expected similarity with eukaryotic lysophospholipases as suggest for esterases from family VI (9). Moreover, EstGtA2 has none of the conserved motifs identified for this family (9). We suggest that EstGtA2 is a member of a new family which diverges from the eight families of the classification by Arpigny and Jaeger (9). It was already suggested that Est30Gs (from

G. stearothermophilus, 1TQH) belonged to a new carboxylesterase family where all members of this new family should have a similar tertiary structure, including the three-helix cap of Est30 (5). EstGtA2 sequence shares 32% identity with Est30Gs, and structural similarity among these two enzymes was suggested on the basis of secondary structure alignment and 3D modelling. The 3D model of EstGtA2 suggests the presence of the three-helix cap structure of Est30 formed by helix B₁', D₁' and D₂'. Despite these similarities, important differences support the hypothesis that EstGtA2 belongs to another family. A high conservation of the hydrophobic residues lining the substrate pocket was also defined as a criterion for members of the new family suggested for Est30Gs (5). However, ~50% of the residues of the hydrophobic specificity pocket are different in EstGtA2 compared to Est30Gs. As shown in Fig. 9, the residue L95(98) (first number corresponds to Est30Gs sequence, while number in brackets correspond to EstGtA2 sequence, i.e. as shown in Fig. 5) included in the pentapeptide GLSLG is replaced by M98 in EstGtA2 (leading to GLSMG). The residues V98(101), F99(102), P118(122), M127(131), V131(135), I171(174), M195(198) and I196(199) in Est30Gs are replaced by T101, L102, A122, G131, G135, M174, V198 and V199 in EstGtA2. Such differences may be responsible for EstGtA2 affinity for longer *p*NP esters (residue substitutions indicated in Fig. 9). EstGtA2 was shown here to be active on C12-C14 and slightly active on C16, which are not hydrolysed by Est30Gs (5).

Another important difference is the optimal temperature of EstGtA2 (50°C) and Est30Gs (70°C) (5, 31). It is generally recognized that salt bridges are involved in protein thermostability since thermophilic protein often display a high number of surface salt bridges in comparison with its mesophilic counterpart (32, 33). Salt bridges were suggested in the EstGtA2 modelled structure and compared with Est30Gs, using ESBRI interface (26) and the following criteria: salt bridges formation was inferred for a pair of oppositely charged (Asp or Glu with Arg, Lys or His) if one pair of Asp or Glu side-chain carboxyl oxygen atoms and side-chain nitrogen atoms of Arg, Lys or His are within a 4 Å distance (34). Analysis of EstGtA2 model (Fig. 10) shows only three salt bridges that met this criterion in comparison with seven for Est30Gs: a first bridge between H61 and E63 of the helix B₁' involved in the formation of the cap structure (3.83 Å); an inter-helices D₁'-D₂' salt bridge (3.66 Å) between R140 and E156, which

seems to maintain these helices in close proximity and stabilize the cap domain; and finally between D196 and H226 (2.93 Å) in the putative catalytic triad.

As shown in Fig. 10, seven pairs of charged residues which met the criteria described above were found in the Est30Gs structure. The salt bridge D193(196)-H223(226) involved in the catalytic triad appears to be equivalent to D196-H226 in EstGtA2. The other six salt bridges are different in Est30Gs when compared with EstGtA2, and should contribute to its higher thermostability. A salt bridge (3.06 Å) links E142 (C terminal end of the helix D₁') to K144 (in the loop just before the N-terminal end of the helix D₂'). In addition, a salt bridge (2.94 Å) formed between K139 and E152 might link the helix D₁' to D₂', contributing to stabilizing the cap domain. It should be noted that K139 (D₁') may also form a salt bridge (3.43 Å) with E59 located in the helix B₁'. Rotamer analysis would be useful for determining which putative salt bridge is most favoured. In one case or another, any salt bridge formed by K139 may contribute to the stabilization of the cap domain by linking the helix D₁' to D₂' or D₁' to B₁'. Moreover, the N-terminal end of the helix D₁' is bound to the N-terminal end of α D by a salt bridge (3.74 Å) formed between E124-K165. These interactions seem to confer a high stability of the cap domain in Est30 compared to EstGtA2 for which the cap structure was stabilized only by the inter-helices D₁'-D₂' salt bridge (R140-E156) and intra-helix B₁' salt bridge (H61-E63). Clearly, these salt bridges and consequent stabilization of the cap seem to be an important determinant of thermostability in carboxylesterase of this family (*G. stearothermophilus*).

Additional salt bridges were found to be potentially stabilizing for Est30Gs structure. An inter helix-strand, α F- β 8 salt bridge formed between K216 and D237 (2.70 Å) might be highly important for the thermostability of Est30Gs since the N-terminal end of the strand β 8 is also involved in a salt bridge (2.78 Å) with N-terminal end of the strand β 7 (R191-E219) that might stabilize the residues D193 and H223 of the catalytic triad located on loops between β 7- α E and β 8- α F respectively. These stabilizing interactions are not observed in the EstGtA2 model since K216 was substituted for a valine (V219) and D237 substituted for an arginine (R240) in EstGtA2 (Fig. 10). In addition of these mutations, R191 is substituted for an aspartate (D194) and E219 for a

histidine (H222) in EstGtA2. The pair D194-H222 formed by these substitutions in EstGtA2 is separated by a distance of 4.74 Å which does not meet the criterion of a 4 Å distance for salt bridge detection. Although an impact on thermostability from such variations in charged residue composition is expected, it remains that their proximity to the catalytic triad might also play a role in the different pH activity curves recorded for the two enzymes compared here (optimal pH for EstGtA2 is 8, and 9 for Est30). The detailed analysis of MSA (Fig. 5) allowed identification of localized sequence features that are probably responsible for the important differences between EstGtA2 and Est30Gs enzymatic properties. Furthermore, these sequence features are well conserved in the cluster N when compared to the sequences related to EstGtA2 (cluster N'), and their mutations in EstGtA2 are also well conserved in the N' family proposed here as shown in Fig. 8.

Analysis of CMB-A2 16S rRNA and EstGtA2 amino acid sequence indicates that CMB-A2 is a novel *G. thermodenitrificans* strain, closely related to strain NG80-2 identified recently in a deep oil reservoir from Northern China (12). The isolation and recovery of CMB-A2 from compost 'grown' in a part of the world devoid of any oil extraction activity (Mauricie region, Quebec) corroborates the hypothesis that NG80-2 originated from a soil environment. NG80-2 would have gained the capacity for alkane oxidization using plasmid pLW1071 in an oil contaminated soil before invading the oil reservoir (13). It will be interesting to investigate the potential ability of CMB-A2 to process longer chain alkanes (i.e. C36) and verify the presence of plasmid similar to pLW1071.

Conclusion

This study allowed identifying a novel carboxylesterase EstGtA2 from *G. thermodenitrificans* which have potential in various industrial applications considering its wide pH range and its thermostability (65°C). This study also suggests the evidence of the existence of a new family of carboxylesterases, tentatively named N and N'. Further work on the correlation between sequence features and enzymatic

properties such as thermostability in EstGtA2 and other esterases and lipases from organisms related to the *Geobacillus* genus is in progress. Our ultimate goal is to develop an understanding of the determinants of properties and stability in this focused group of enzymes.

Acknowledgments

D.M.C. acknowledges a scholarship from the NSERC.

Funding

Natural Sciences and Engineering Research Council of Canada (NSERC).

References

1. Okuda, H. (1991) Esterase in A Study of Enzymes (Kuby, S.A., ed.) Vol. 1, pp. 563–577, CRC press, Boca Raton, Florida
2. Verger, R. (1997) Interfacial activation of lipases: facts and artifacts. *Trends Biotechnol.* 15, 32–38
3. Jaeger, K-E., Ransac, S., Dijkstra, B.W., Colson, C., Heuvel, M., and van Misset, O. (1994) Bacterial lipases. *FEMS Microbiol. Rev.* 15, 29–63
4. Liu, P., Wang, Y.F., Ewis, H.E., Abdelal, A., Lu, C.D., and Weber, I.T. (2003) Crystallization and preliminary X-ray diffraction data for the carboxylesterase Est30 from *Bacillus stearotherophilus*. *Acta. Cryst. D* 59, 1472–1473
5. Liu, P., Wang, Y.F., Ewis, H.E., Abdelal, A., Lu, C.D., Harrison, R.W., and Weber, I.T. (2004) Covalent reaction intermediate revealed in crystal structure of the *Geobacillus stearotherophilus* carboxylesterase Est30. *J. Mol. Biol.* 342, 551–561
6. Bornscheuer, U.T. (2002) Microbial carboxyl esterases: classification, properties and applications in biocatalysis. *FEMS Microbiol. Rev.* 26, 73–81
7. Gupta, M.N. and Roy, I. (2004) Enzymes in organic media. Forms, functions and applications. *Eur. J. Biochem.* 271, 2575–2583
8. Ollis, D.L., Cheach, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S.M., Harel, M., Remington, S.J., Silman, I., Schrag, J., Sussman, J.L., Verschueren, K.H.G., and Goldman, A. (1992) The α/β hydrolase fold. *Protein Engin.* 5, 197–211
9. Arpigny, J.L. and Jaeger, K-H. (1999) Bacterial lipolytic enzymes: classification and properties. *Biochem. J.* 343, 177–183
10. Wilson, K. (1987) Preparation of genomic DNA from bacteria in *Current protocol in Molecular Biology* (Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Smith, J. A., Seidman, J. G., and Struhl, K., eds.), pp. 2.4.1–2.4.5, Wiley, New York
11. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A laboratory Manual*, p. 134, Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York

12. Wang, L., Tang, Y., Wang, S., Liu, R-L., Liu, M-Z., Zhang, Y., Liang, F-L., and Feng, L. (2006) Isolation and characterization of a novel thermophilic *Bacillus* strain degrading long-chain n-alkanes. *Extremophiles* 10, 347–356
13. Feng, L., Wang, W., Cheng, J., Ren, Y., Zhao, G., Gao, C., Tang, Y., Liu, X., Han, W., Peng, X., Liu, R., and Wang, L. (2006) Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir. *Proc. Natl Acad. Sci. USA* 104, 5602–5607
14. Johnson, M., Zaretskaya, I., Raytselis, Y., Merezuk, Y., McGinnis, S., and Madden, T.M. (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res.* 36, W5–W9
15. Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. (2007) GenBank. *Nucleic Acids Res.* 36, D25–D30
16. Altschul, S.F., Madden, T.L., Scha" ffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402
17. Marchler-Bauer, A., Panchenko, A.R., Shoemaker, B.A., Thiessen, P.A., Geer, L.Y., and Bryant, S.H. (2002) CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res.* 30, 281–283
18. Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596–1599
19. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680
20. Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425
21. Swofford, D.L. and Sullivan, J. (2002) Phylogeny inference based on parsimony and other methods using PAUP in *The Phylogenetic Handbook, A Practical Approach to DNA and Protein Phylogeny* (Salemi, M. and Vandamme, A-M., eds.) Vol. 1, pp. 160–202, Cambridge University press, Cambridge
22. Lambert, C., Leonard, N., De Bolle, X., and Depiereux, E. (2002) ESyPred3D: Prediction of protein 3D structures. *Bioinformatics* 18, 1250–1256

23. Hooft, R.W.W., Vriend, G., Sander, C., and Abola, E.E. (1996) Errors in protein structures. *Nature* 381, 272
24. Laskowski, R.A., MacArthur, M.W., Moss, D., and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26, 283–291
25. Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22, 195–201
26. Constantini, S., Colonna, G., and Facchianno, A.M. (2008) ESBRI: a web interface for evaluating salt bridges in proteins. *Bioinformation* 1, 137–138
27. Smith, P.K., Krohn, R.I., Hermanson, G.T., Mailla, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J., and Klenk, D.C. (1985) Measurement of protein using bicinchoninic acid. *Anal. Biochem.* 150, 76–85
28. Imamura, S. and Kitaura, S. (2000) Purification and characterization of a monoacylglycerol lipase from the moderately thermophilic *Bacillus* sp. H-257. *J. Biochem.* 127, 419–425
29. Kitaura, S., Suzuki, K., and Imamura, S. (2001) Monoacylglycerol lipase from moderately thermophilic *Bacillus* sp. strain H-257: molecular cloning, sequencing and expression in *Escherichia coli* of the gene. *J. Biochem.* 129, 397–402
30. Montoro-Garcia, S., Martinez-Martinez, I., Nevarro- Fernade' z, J., Takami, H., Garcia-Carmona, F., and Sanchez-Ferrer, A. (2009) Characterization of a novel thermostable carboxylesterase from *Geobacillus kaustophilus* HTA426 shows the existence of a new carboxylesterase family. *J. Bacteriol.* 191, 3076–085
31. Ewis, H.E., Abdelal, A.T., and Lu, C.D. (2004) Molecular cloning and characterization of two thermostable carboxyl esterases from *Geobacillus stearothermophilus*. *GENE* 329, 187–195
32. Takano, K., Tsuchimori, K., Yamagata, Y., and Yutani, K. (2000) Contribution of salt bridges near the surface of a protein to the conformational stability. *Biochemistry* 39, 12375–12381
33. Strop, P. and Mayo, S.L. (2000) Contribution of salt bridges to protein stability. *Biochemistry* 39, 1251–1255
34. Kumar, S. and Nussinov, R. (1999) Salt bridge stability in monomeric proteins. *J. Mol. Biol.* 293, 1241–1255

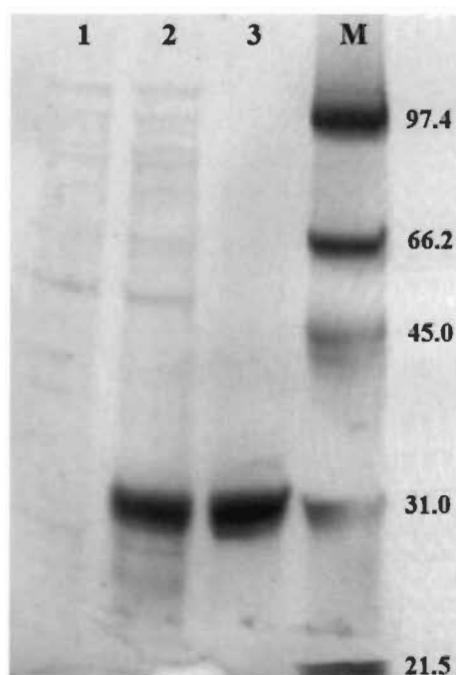


Figure 1 SDS-PAGE of the recombinant EstGtA2. Lane 1: *E. coli* M15 [pREP4]/pQE31-estGtA2 total proteins extract before induction of EstGtA2; lane 2: total protein extract after induction 2.5 h with 1 mM IPTG at 37°C; lane 3: purified His₆-tagged EstGtA2; M: low range molecular weight marker (BioRad) bands are identified with respective MW: 97.4, 66.2, 45, 31 and 21.5 kDa).

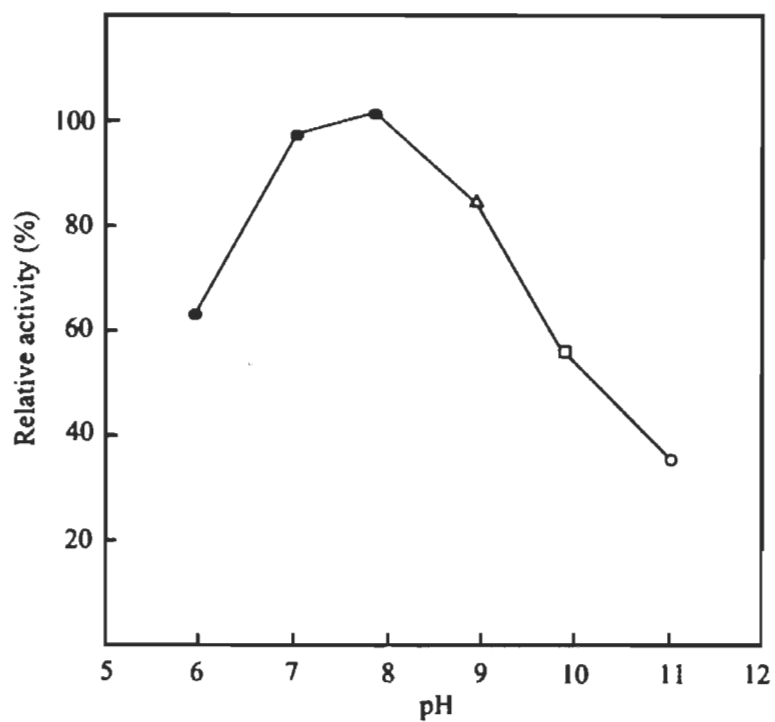


Figure 2 Activity of EstGtA2 at different pH values. Activity of EstGtA2 was assayed with 14 μ g of protein and 50 mM *p*NP caprylate at 50°C in appropriate buffers. Closed circles: 50 mM sodium phosphate, pH 6–8; open triangle: 50 mM Tris–HCl, pH 9; open squares: 50 mM CAPS, pH 10; open circles: sodium phosphate dibasic, pH 11. Activity at pH 8 was defined as 100% relative activity.

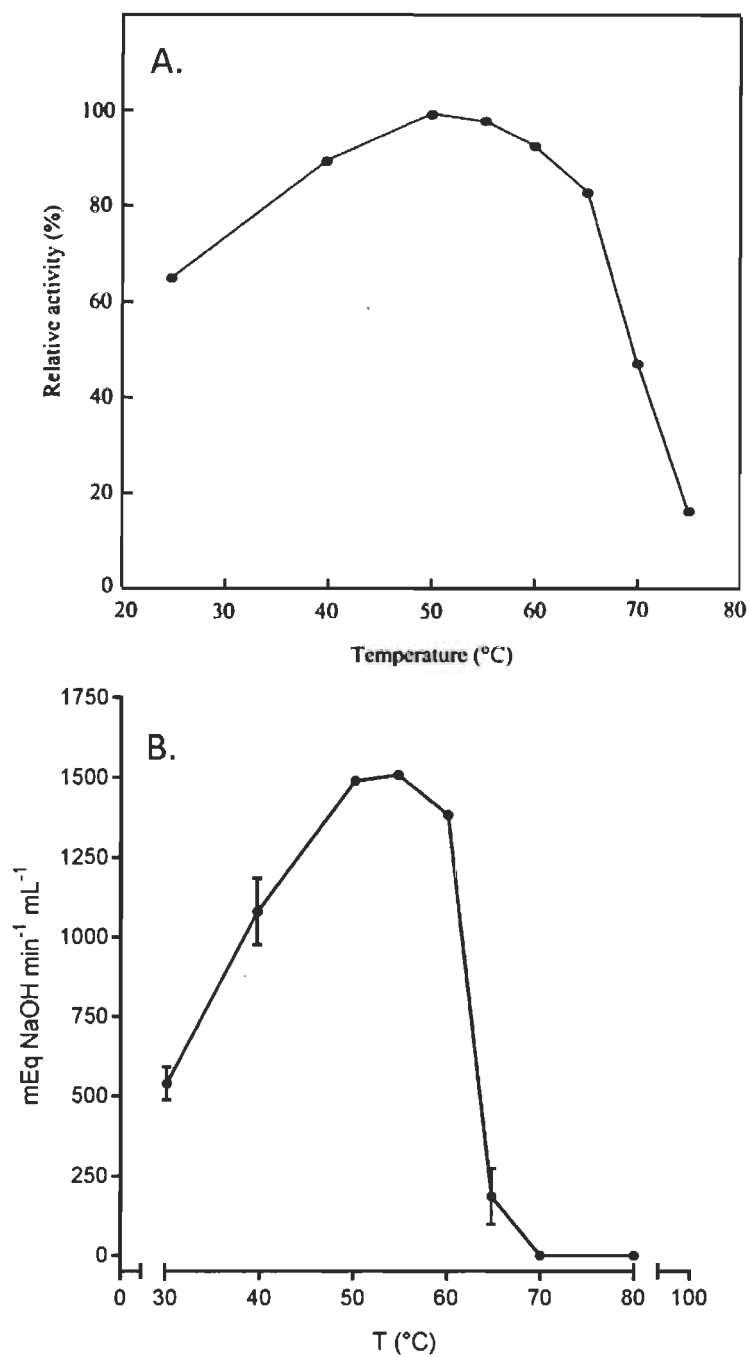


Figure 3 Activity of EstGtA2 at different temperatures. A. Activity of EstGtA2 was assayed with 14 μ g of protein and 50 mM *p*NP caprylate in 50 mM phosphate buffer, pH 8 at specified temperatures. Activity at 50°C was defined as 100% relative activity. B. Hydrolysis of tributyrin catalyzed by EstGtA2, measured from 30 to 80°C by pH-stat titrimetry. Hydrolysis is reported in milliequivalents of NaOH per min by mL of solution.

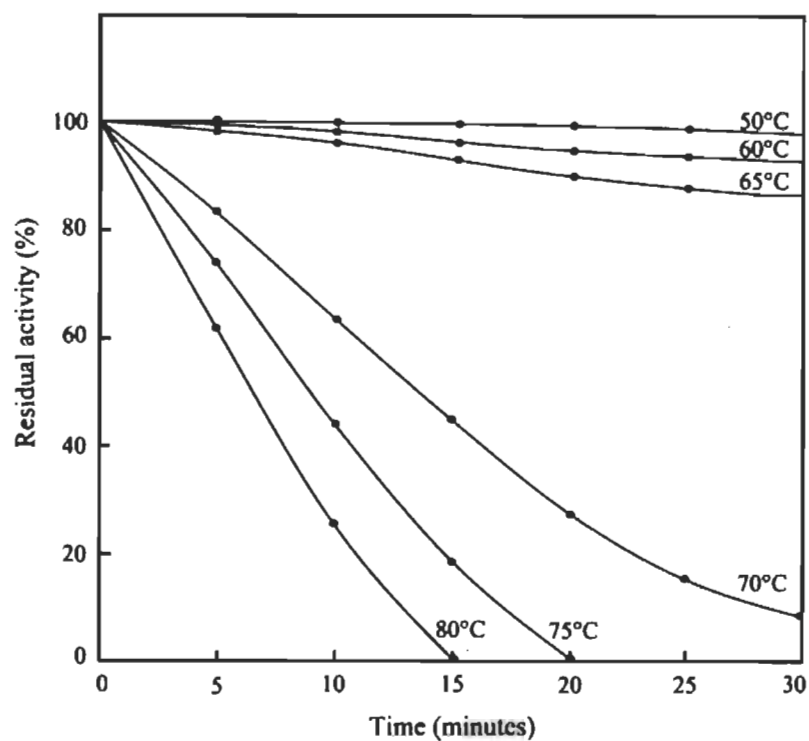


Figure 4 **Stability of EstGtA2 at different temperatures.** Thermal stability was assayed at pH 8. Enzyme was incubated for different period of time at 50, 60, 65, 70, 75 and 80°C and then activity was measured under the standard reaction conditions at 50°C.

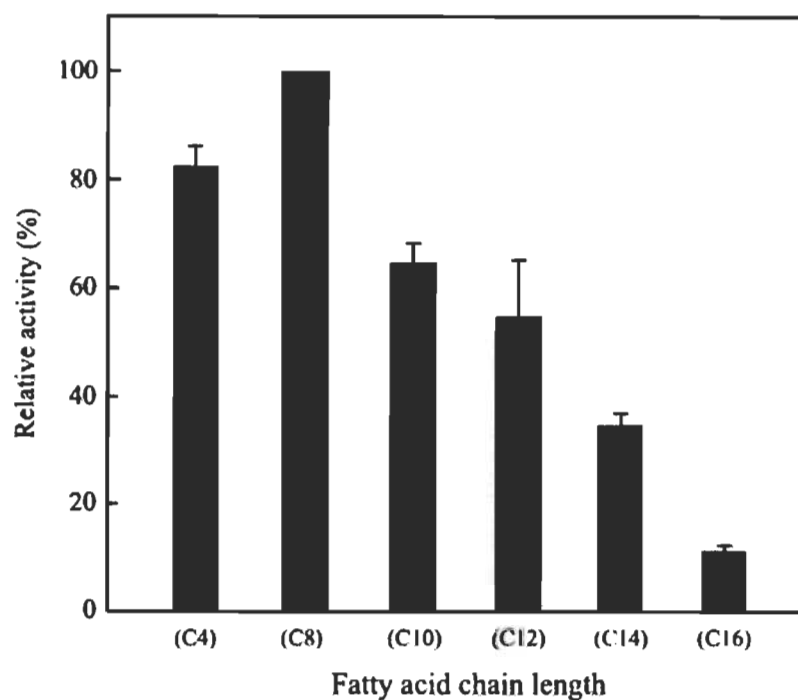


Figure 5 Activity of EstGtA2 toward *p*NP esters of different fatty acids chain length. Activity of EstGtA2 was assayed with 14 μ g of protein on specified *p*NP esters (50 mM) at 50°C in 50 mM phosphate buffer pH 8. Activity using *p*NP caprylate (C8) as substrate was used as 100% relative activity.

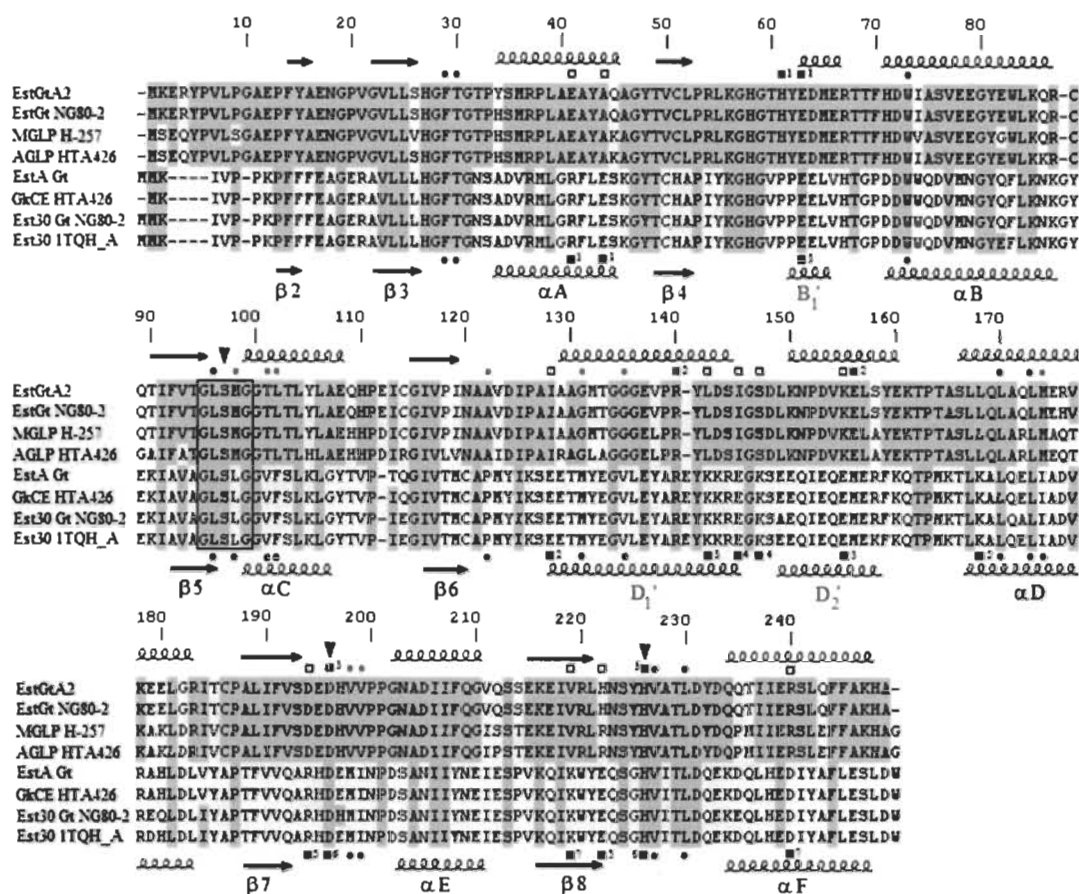


Figure 6 Multiple sequence alignment for EstGtA2 from *G. thermodenitrificans* and related *Geobacillus* esterases. Sequences are grouped according to similarity. Closest homologues of EstGtA2 found in BLAST analysis (first three sequences) were EstGt from *G. thermodenitrificans* NG80-2, MGLP from *Bacillus* sp. H-257, AGLP from *G. kaustophilus* HTA426. Sequences related to Est30Gs from *G. stearothermophilus* were EstA from *G. thermoleovorans*, GkCE from *G. kaustophilus* HTA426 and Est30 from *G. thermodenitrificans*. Conserved residues are embedded in a solid background. Secondary structures of Est30Gs (as calculated from X-ray resolved structure) are indicated below the MSA, and secondary structures modelled for EstGtA2 are indicated above the MSA. Amino acids of the catalytic triad are identified with arrows. The putative salt bridges in the EstGtA2 and Est30Gs structures were analysed with ESBRI. Amino acids involved in salt bridges are identified by closed black squares and annotated with respective numbers in pairs (1-3 for EstGtA2 and 1-7 for Est30Gs) above the MSA for EstGtA2, and underneath for Est30Gs. Open squares above the MSA represent potential loss of salt bridges in EstGtA2 compared with Est30Gs. The residues of the hydrophobic specificity pocket identical between Est30Gs and EstGtA2 are indicated with black closed circles. Grey closed circles above the MSA represent differences in EstGtA2 specificity pocket compared with Est30Gs. MSA positions are numbered using EstGtA2 amino acid sequence.

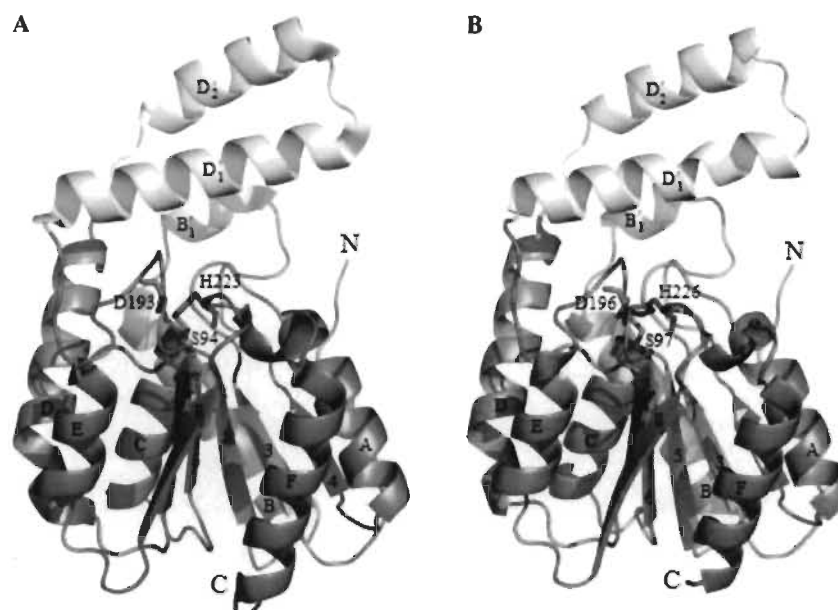


Figure 7 **3D model of EstGtA2.** Ribbon structure of Est30Gs (A) and modelled structure of EstGtA2 (B). Amino acids of the catalytic triad are identified and shown in the centre of the respective structure. Alpha helices are identified (from A to F) and beta strands (2-8). The lid helices are identified (D1', D2' and B1') and are protruding above the enzyme, and shown light grey.

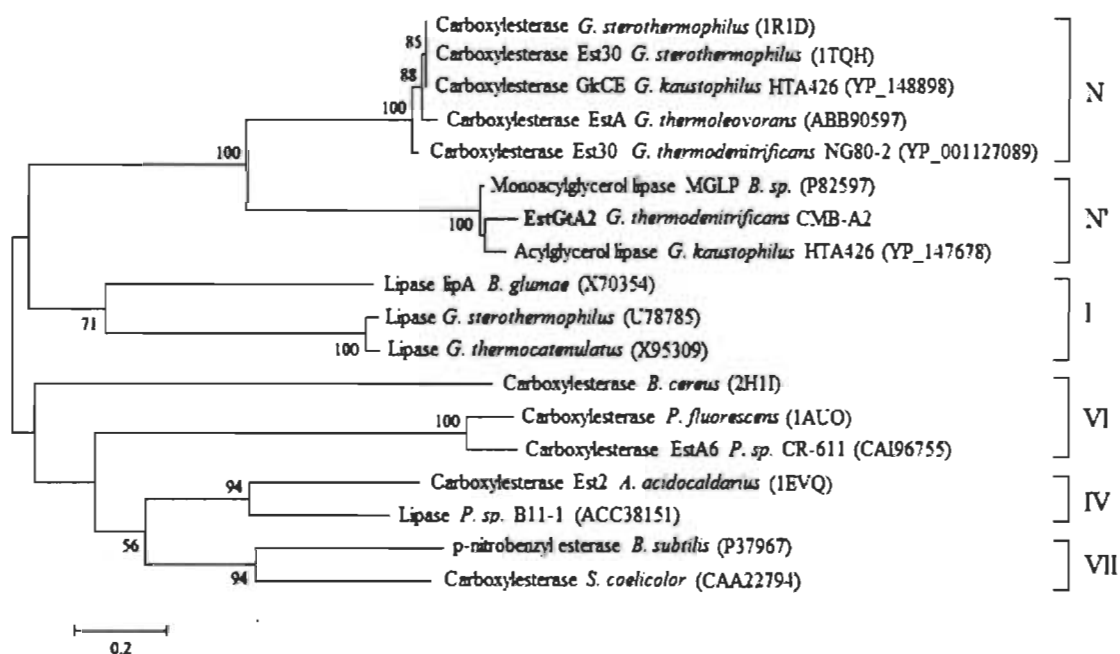


Figure 8 **Phylogenetic tree of EstGtA2 and related lipolytic enzymes.** Phylogenetic tree representing EstGtA2 with related homologues and with lipolytic enzymes used in the classification of Arpigny and Jaeger (9). Roman numerals represent the lipolytic enzyme family according to the original classification. N and N' represent the new clusters of carboxylesterases as suggested here.

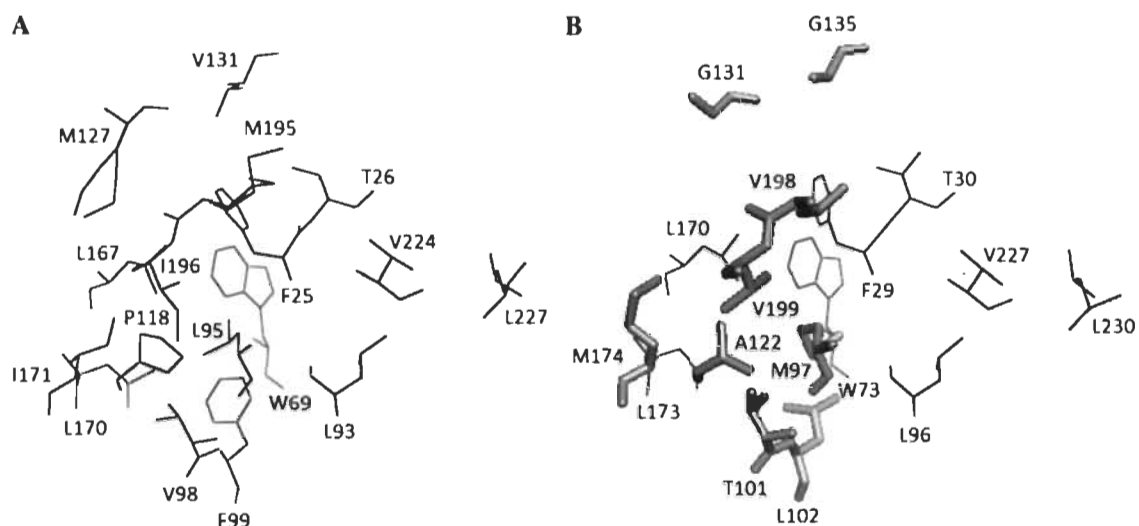


Figure 9 **Hydrophobic specificity pocket of Est30Gs and EstGtA2.** Residues of the hydrophobic specificity pocket for substrate of Est30Gs (A), and as predicted by EstGtA2 modelling (B), are shown. Amino acid residue substitutions for EstGtA2 are shown as thick sticks.

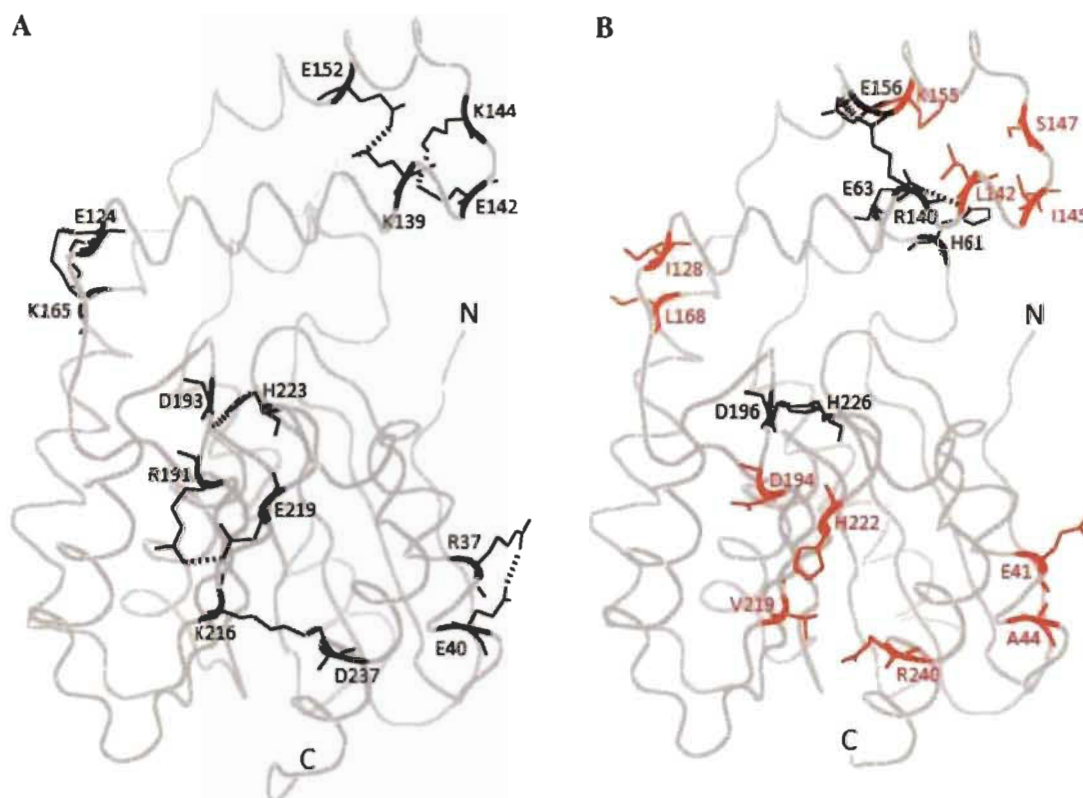


Figure 10 Prediction of salt bridges in EstGtA2. Predicted salt bridges in Est30 (A) and EstGtA2 (B) structures are shown by dotted lines. Putative salt bridge residues ($< 3 \text{ \AA}$) are shown in black, while amino acid substitutions leading to loss of bridges in EstGtA2 (compared to Est30Gs) are shown in red.

CHAPITRE IV

N-TERMINAL PURIFICATION TAG ALTERS THERMAL STABILITY OF THE CARBOXYLESTERASE ESTGTA2 FROM *G. THERMODENITRIFICANS* BY IMPAIRING REVERSIBILITY OF THERMAL UNFOLDING

Ce chapitre a fait l'objet d'une publication dans le journal scientifique « Protein and Peptide Letters » hébergé par *Bentham Science publishing group*.

David M. Charbonneau, Fatma Meddeb-Mouelhi and Marc Beauregard. 2012. N-terminal purification tag alters thermal stability of EstGtA2 by impairing reversibility of thermal unfolding. *Prot. Pep. Lett.* 19, 264-269

<http://www.eurekaselect.com/89639/article>

4.1 Contribution des auteurs

Ces recherches et l'ensemble des expériences qui ont mené à la publication de cet article ont été conçus par David Charbonneau, Fatma Meddeb et Marc Beauregard qui ont aussi participé à l'analyse des résultats. David Charbonneau a rédigé en entier le manuscrit qui a été révisé par Fatma Meddeb et Marc Beauregard avant sa publication.

4.2 Résumé de l'article

La nouvelle carboxylesterase thermostable EstGtA2 de *Geobacillus thermodenitrificans* (Accession AEN92268) a été surexprimée de façon soluble et fonctionnelle chez *E.coli* de façon recombinante avec la fusion d'un peptide (étiquette) de purification en N-terminal (His₆-tag) et purifiée par chromatographie d'affinité sur métal immobilisé (IMAC) en utilisant une résine de nickel. Nous avons récemment rapporté les propriétés générales par une caractérisation initiale de EstGtA2 (forme

recombinante) qui a été purifiée à l'aide de l'étiquette de purification (His₆-tag) en N-terminal. Dans ce présent rapport, nous évaluons l'impact de cette étiquette en N-terminal sur la thermostabilité de EstGtA2 en comparant la stabilité de EstGtA2-tagged (avec son étiquette de purification) et EstGtA2-cleaved (sans son étiquette de purification). Suite à la purification IMAC, l'étiquette a été clivée par l'hydrolyse protéolytique de la thrombine. La stabilité conformationnelle de EstGtA2 a été étudiée par dichroïsme circulaire et par la mesure de son activité enzymatique à l'aide d'une méthode de détection colorimétrique. Les résultats montrent que l'étiquette de purification n'a pas d'effet indésirable sur l'activité de EstGtA2 et sur la température de fusion. Par contre, après dénaturation thermique, la forme étiquetée de EstGtA2 ne récupère pas le même niveau d'activité que la forme dépourvue de cette même étiquette. De plus le suivi du repliement conformationnelle durant la dénaturation (dénaturation thermique), ainsi durant la renaturation (retour à basse température) montre que la forme clivée de EstGtA2 récupère environ 83 % de sa conformation initiale alors que la forme étiquetée ne récupère pas cette même conformation, ce qui suggère que l'étiquette fusionnée en N-terminal, très utile pour la purification rapide de EstGtA2, compromet la réversibilité de dénaturation thermique. Quoique l'étiquette de purification n'a pas d'effet indésirable sur l'activité de EstGtA2 et sur son T_m , celle-ci compromet grandement sa thermostabilité, une propriété qui implique non seulement la rigidité structurale, mais aussi sa réversibilité après dénaturation thermique. Les résultats montrent l'importance de retirer cette étiquette pour l'étude de la thermostabilité des enzymes en général.

4.3 Troisième article scientifique

N-terminal purification tag alters thermal stability of the carboxylesterase EstGtA2 from *G. thermodenitrificans* by impairing reversibility of thermal unfolding

David M. Charbonneau^{1,2}, Fatma Meddeb-Mouelhi^{1,3} and Marc Beauregard^{1,2,*}

1. Département de chimie-biologie, Université du Québec à Trois-Rivières, 3351 Boul. Des Forges, C.P. 500 Trois-Rivières (Québec) G9A 5H7,
2. PROTEO, Université Laval, 2705 Boul. Laurier, Ste-Foy (Québec) G1V 4G2
3. Buckman Laboratories of Canada, 351, Joseph-Carrier Vaudreuil-Dorion (Québec) J7V 5V5, Canada

*Corresponding author : Marc Beauregard, Département de chimie-biologie, Université du Québec à Trois-Rivières, C.P. 500, Trois-Rivières (Québec), G9A 5H7, Canada, Tel: 819 376 5011 (Ext. 3354), Fax: 819 376 5084, email: marc.beauregard@uqtr.ca

Keywords: Carboxylesterase, thermal stability, thermal unfolding, tag peptide, circular dichroism, *Geobacillus thermodenitrificans*

Published in Protein & Peptide Letters

Protein & Peptide Letters 2012, 19, 264-269

Article available online at: <http://www.eurekaselect.com/89639/article>

Abstract

The novel thermostable carboxylesterase EstGtA2 from *G. thermodenitrificans* (accession no. AEN92268) was functionally expressed and purified using an N-terminal fusion tag peptide. We recently reported general properties of the recombinant enzyme. Here we report preliminary data on thermal stability of EstGtA2 and of its tagged form. Conformational stability was investigated using circular dichroism and correlated with residual activity measurements using a colorimetric assay. The tag peptide had no considerable impact on the apparent melting temperature: T_m value = 64.8°C (tagged) and 65.7°C (cleaved) at pH 8. After thermal unfolding, the tag-free enzyme rapidly recovered initial activity at 25°C (1.2 U mg⁻¹), which was corroborated by substantial refolding (83%) as determined by far-UV CD transitions. However, after thermal unfolding, the purification tag drastically decreased specific activity at 25°C (0.07 U mg⁻¹). This was corroborated by the absence of refolding transition. Although the purification tag has no undesirable impact on activity before thermal unfolding as well as on T_m , it drastically hinders EstGtA2 refolding resulting in a major loss of thermal stability.

Introduction

The expression and recovery of recombinant proteins in a highly purified and well-characterized form has become a major task in biotechnology. The addition of an affinity tag peptide in N- or C-terminal of a recombinant protein is now a widely used strategy for its production and rapid purification. Affinity tagging also provides a powerful tool for high throughput purification of proteins and/or mutant generation and selection. Different tagging systems have been developed in recent years [1, 2]. One of the most common strategies consists in fusing a target protein with a sequence encoding a polyhistidine tag for one-step purification using immobilized nickel affinity chromatography. Polyhistidine tags (6 x or 10 x his-tag) usually vary from 6 to more than 20 amino acids in length depending of the chosen system and endonuclease restriction sites used for cloning. More than 90% of the X-ray crystal structures solved

and deposited in PDB is based on recombinant material and among them, 60% was purified using some sort of His-tag [3,4]. In most cases, the tag is unresolved in the crystal structure indicating that it does not adopt a unique or dominant conformation, which further suggest that it maintains limited interaction with the rest of the folded protein. So, it is assumed that purification tags have generally no significant impact on a resolved crystal structure. However, other functional parameters might be more sensitive to the presence of an exogenous tag peptide. It was reported that a His-tag may affect substrate specificity, enantioselectivity, protein solubility and oligomerization [5-8]. In some cases it can hinder crystal nucleation and/or growth. Protein thermostability is a multicomponent property. The measure of thermal stability of a protein often simply refers to its melting temperature (T_m). However, ability to refold after thermal unfolding is an important component of thermal stability for protein from thermophiles. In many cases, thermal stability data are reported for tagged recombinant proteins and the potential impact of purification tags on thermostability has not been addressed in relevant studies [9-12]. We recently reported a novel thermostable carboxylesterase, EstGtA2 from *Geobacillus thermodenitrificans* (GenBank access no. JN031579), and analyzed general properties of the recombinant protein [13]. The protein was produced and purified using an N-terminal fusion peptide containing a His-tag. The tag peptide had no considerable effect on general properties of the recombinant EstGtA2, such as optimal temperature and pH for activity. However, thermal stability data are quite different. Protein thermal stability involves two distinct components: Rigidity (proportional with T_m) and refoldability after thermal unfolding. Thermal stability experiments are usually performed as follows: incubation of a protein at a specified temperature for a given period of time, followed by activity measurement at lower temperature (often at room temperature). When the incubation temperature is higher than the T_m , then activity detection at room temperature becomes dependant of protein refolding. We wanted to address the potential impact of the N-terminal His-tag peptide on conformational reversibility of EstGtA2 after thermal unfolding in order to explain the drastic decrease in thermal stability for the tagged protein in comparison with the tag-free EstGtA2 (cleaved).

Materials and methods

Buffers and chemicals

All chemicals were purchased from Sigma-Aldrich. Thermal denaturation experiments were performed in the respective buffers: 20 mM sodium citrate/citric acid (pH 3, 4 and 5), sodium phosphate (pH 6, 7 and 8), and CAPS (pH 9, 10 and 11). All buffers were prepared with distilled/ deionised water and filter-sterilized (0.22 μ m).

DNA cloning and protein expression

The novel thermostable EstGtA2 was isolated from *G. thermodenitrificans* CMB-A2 (accession no. GQ293454) and the pQE31-EstGtA2 construct (Charbonneau *et al.*, 2010) was used as template for PCR. The EstGtA2 ORF was amplified by PCR using the following primers pair: Forward primer (*NdeI*) 5'-CTCGGTACATATGAAAGAACGATATCCTGTACTIC-3' and Reverse primer (*EagI*) 5'-GTGCGGCCGTCAAGCATGTITG GCGAAAAACTG-3'. It was then cloned in *E. coli* BL21 Gold using the pET28 expression vector using *NdeI* and *EagI* restriction site. The recombinant protein was functionally expressed by induction at an $OD_{600\text{ nm}} = 0.6$ with 0.5 mM IPTG for 16h at 25°C.

Protein purification and cleavage of the N-terminal tag

The recombinant protein was purified under native conditions using immobilized nickel affinity chromatography (Ni-NTA). Half of the recombinant protein was dialyzed directly in 20 mM sodium phosphate pH 8 for 48h at 4°C using 10,000 MWCO dialysis membranes (EstGtA2tagged). The other half was dialyzed against 50 mM Tris-HCl pH 8 and further dialyzed in thrombin buffer (50 mM Tris-HCl, 150 mM NaCl and 2.5 mM CaCl₂). The thrombin was then added (10 NIH Units) at ratio thrombin: EstGtA2 = 1: 5000 and incubated for 24h at 25°C under gentle rotation (hybridization incubator). The cleaved protein (EstGtA2_{cleaved}) was purified from the free His tag and from non-cleaved protein using Ni-NTA chromatography by recuperating the flow through.

The protein was then dialyzed in Tris-HCl 50 mM, pH 8 and buffer exchange was performed by further dialysis of the purified cleaved protein in 20 mM sodium phosphate pH 8. The protein concentration for the tagged and cleaved protein was determined after dialysis by the bicinchoninic acid method [14] and the purity of each protein samples was verified by SDS-PAGE.

Enzyme activity assay

Activity measurements were performed in a final volume of 0.2 ml in 96-wells microplates. Reaction (0.2 mL) contained 50 or 100 μ M *p*NP caprylate in 20 mM sodium phosphate pH 8. For kinetics experiments Fig. (2), protein samples (0.05 mg/ml) were incubated at 80°C. A volume of 0.1 mL (5 μ g of protein) was assessed for residual activity measurements. Enzyme activity was determined by measuring the amount of *p*NP released after hydrolysis of *p*NP caprylate. The amount of *p*NP released was recorded by monitoring the absorbance at 405 nm. A standard curve for *p*-nitrophenol absorbance (at pH 8 and 25°C) as function of concentration (μ M) allowed determining the rate of hydrolysis and enzyme specific activity. One unit of enzyme activity was defined as the amount of *p*-nitrophenol released by minute per microgram of protein. For the end point experiment Fig. (2), protein samples (0.1 mg/mL) were incubated at 80°C. A volume of 0.1 mL (10 μ g of protein) was assessed for residual activity (after 10 min reaction). Relative activity was reported in percentage (%) from the initial activity (activity at 25°C), measured before incubation at 80°C.

CD measurements

CD spectra were recorded on a Jasco-720 spectropolarimeter. Far-UV (190-250 nm) CD spectra were recorded in a 0.1 cm path length quartz cuvette with protein concentration of 0.4 mg/mL in 20 mM sodium phosphate buffer pH 8.0. Spectra were accumulated from 30 to 80°C (at each 5°C) with a scan speed of 20 nm per minute and data were collected at 222 nm for each temperature. Temperature of the cuvette was controlled using a Neslab RTE-III circulating water bath connected to the water-jacketed quartz cuvettes.

Conformational and thermodynamics (unfolding/refolding)

Direct assessment of protein refolding was investigated by monitoring secondary structures. Thermal unfolding/ refolding profiles were derived from far-UV circular dichroism (CD) measurements. For CD experiments, protein samples were adjusted at 0.4 mg/mL in 20 mM sodium phosphate pH 8 for the tagged and cleaved EstGtA2 respectively. CD spectra were measured from 190-250 nm as the temperature of the cuvette was increased from 30 to 80°C, transition (F → U) at a rate of 0.8°C/min and the protein samples were cooled from 80 to 30°C, transition (U → R), at a rate of 0.8°C/min. CD data were converted in molar ellipticity by taking into account the difference in molecular mass ($\Delta MW = 2$ kDa) between the tagged and cleaved protein. CD data were also converted in molar CD ($\Delta\epsilon$) and the secondary structures for the folded, unfolded and refolded proteins were calculated using SELCON and CDSSTR softwares which calculate different assignments of secondary structures on the basis of CD spectra measured from different proteins for which high quality X-ray diffraction data are available [15-17]. The databank CLSTR representing 48 soluble/denatured proteins was used for determining changes in secondary structures during unfolding and refolding. Molar ellipticity at 222 nm recorded for each temperature was used as a conformational probe for helices. Thermal unfolding curves were fitted on a two-state model (F ↔ U). Extrapolation and least-square analysis from the pre- and post-transition regions for thermal unfolding profiles from 25-95°C allowed determination of the equations for y_F and y_U . The equilibrium constant was determined at each temperature using equation 1.

$$\text{eq 1. } K = f_U / (1 - f_U) = f_U / f_F = (y_F - y) / (y - y_U)$$

The free energy associated with the thermal unfolding was calculated at each temperature by the equation 2.

$$\text{eq2. } \Delta G = -RT \ln K = -RT \ln [(y_F - y) / (y - y_U)]$$

The apparent T_m values were then determined by plotting ΔG as function of temperature, where $T_m = T$ at $\Delta G = 0$.

Results

Expression and purification of the recombinant EstGtA2

The protein was cloned and functionally expressed in *E. coli* BL21 Gold using the pET28 system at a yield of 60 mg/L. The ORF was cloned using the *NdeI* and *EagI* restriction sites which results in a fusion peptide of 20 amino acids (MGSSHHHHHSSGLVPR|GSH) added at the N-terminal end of the translated recombinant protein which allows easy one-step purification using Ni-NTA. After purification, we then removed the N-terminal tag using the thrombin protease (cleavage site underlined in the peptide sequence shown above) which result in a reduced apparent molecular mass of 2 kDa, (see Fig. 1). We then evaluated the impact of the His-tag peptide on the stability of EstGtA2 by comparing activity, thermostability and conformational reversibility of the tagged protein and after removal of the His-tag, the latter protein is hereafter named “cleaved protein”.

Thermal stability of EstGtA2-tagged and -cleaved

Activity and thermal stability was investigated by measuring the catalytic rate of hydrolysis for the tagged and cleaved protein at 25°C (without previous incubation) and after incubation at 80°C for different period of time Fig. 2 (A and B). The specific activity at 25°C was calculated in the transition region (between 15 and 60 sec). The enzyme activity was identical ($1.2 \text{ U} \cdot \text{mg}^{-1}$) for the tagged and cleaved EstGtA2. The initial rate from 0-15 sec was considerably lower for heated samples; however after the first 15 sec, the cleaved protein recovers the same rate than unheated samples (absorbance change per unit of time). The time lag observed for the cleaved EstGtA2 might come from initial protein refolding. On the other hand, the rate observed for the heated tagged enzyme was drastically decreased and retained only 6% of initial activity ($0.07 \text{ U} \cdot \text{mg}^{-1}$), see Fig. 2B. This indicates that the tagged protein did not refold during the 60 sec measurement period. Thermal stability was further investigated by incubation of protein samples at 80°C for different periods followed by measurement of residual activity after a fixed reaction time period of 10 min at 25°C. The amount of

p-nitrophenol released after hydrolysis of substrate (after 10 min reaction) was taken as end point to estimate residual activity at 25°C after incubation at 80°C Fig. (3). Residual hydrolysis of *p*NP caprylate was measured in percentage, taking the amount of *p*NP released at 25°C (without incubation) as 100% activity. Initial activity was identical for the tagged and cleaved protein. The cleaved EstGtA2 retain 100% of its initial activity after 180 min incubation at 80°C Fig. (3), which is in line with the rapid recovery of specific activity discussed above Fig. (2). The cleaved EstGtA2 finally lost its activity (measured within the 10 min reaction period at 25°C) after 240 min incubation at 80°C. For the tagged protein, about 50% of residual activity was detected after 30 min incubation at 80°C, and complete inactivation was observed after 60 min (an incubation period 4-fold shorter than for the cleaved protein (240 min)). This result demonstrates that the fusion peptide had a major impact on EstGtA2 thermostability. In order to understand how the tag peptide affects protein thermostability, we investigated conformational changes during thermal unfolding and refolding.

Conformational and thermodynamic analysis (unfolding/refolding)

The CD spectra recorded at 30°C (folded) and 80°C (unfolded) were very similar for the tagged and cleaved proteins and typical for proteins with mixed alpha/beta content Fig. (4). At high temperature, both proteins show a similar organization of the unfolded state, but after cooling, the CD spectra for the cleaved and tagged protein were quite different. The helical content decreased progressively during thermal unfolding (22 to 7% and from 24 to 14% for the cleaved and tagged EstGtA2 respectively) leading to an increase in unordered structures (28 to 45 and 37 to 57%). After cooling to 30°C, EstGtA2_{cleaved} recovered about 95% (21/22%) of the initial helices while EstGtA2_{tagged} recovered only 64% (18/28%) of initial helical content. The cleaved protein also recovered the initial value for unordered structures while the tagged protein maintained high content in unordered structures reached during unfolding Fig. (4). Molar ellipticity at 222 nm recorded for each temperature was used as a conformational probe for helices Fig. (5). Partial reversibility was observed below 85°C and the unfolding/refolding experiments were performed from 30 to 80°C. Extrapolation and least-square analysis

from the pre- and post-transition regions for thermal unfolding profiles from 25-95°C allowed determination of the equations for y_F and y_U . The apparent T_m values calculated for the cleaved and tagged protein were 65.7 and 64.8°C respectively. The tag peptide had but little impact on the melting temperature and protein unfolding, while it clearly inhibited the refolding rate and modified the organization of the refolded state (R), see Fig. (4). The unfolding/refolding profile pattern is highly cooperative for EstGtA2cleaved whereas the EstGtA2 tagged did not show any refolding transition Fig. (5). The refolded tagged protein seems to be trapped in a misfolded state. It is clear that the effect of the tag peptide has a destabilizing impact on the protein refolding kinetics and cooperativity. This further supports the hypothesis that the major difference in thermal stability (residual activity detected after incubation at 80°C) does not come from a tag-induced destabilization (unfolding or lower T_m) but rather to an alteration of the refolding component.

Conclusion

The widely used His-tag purification strategy has been shown to impair refolding of the carboxylesterase EstGtA2. Potential consequences on fundamental as well as on applied features are to be considered when using such a tag. Regarding industrial use of such tagged enzymes, the tag may jeopardize enzyme recycling (i.e. reuse in repetitive treatments) or decrease enzyme lifetime. It can also be responsible for underestimation of protein thermostability and causes misinterpretation in comparing thermal stability data reported for homologous structures (tagged vs. untagged). Carboxylesterase and lipases have several potential applications and the thermostability is often a major factor determining its viability as catalyst. Usually, thermal stability is assessed by measuring activity after cooling from a heat stress. The “thermal stability” measured in this way is therefore dependant on the ability of the thermally denaturated protein to refold. Often, the potential impact of a His-tag on such measurement has not been addressed in relevant studies. Refolding may also be an issue for recovery of proteins from inclusion bodies in high yield recombinant prokaryotes. In a recent study (Wear *et al.*, 2006) authors have addressed the question of *in vitro* refolding for the protein FKBP-12, after

its chemical denaturation [19]. On the basis of their results, it was suggested that the His-tag reduced thermostability by inducing the formation of misfolded conformer(s) with reduced activity. Such misfolded state(s) might be corrected *in vivo* with the help of chaperones. FKBP-12 has no relatedness to EstGtA2 (18% residues identity) and adopt a different fold. Taken together these observations suggest that the impact of a purification tag in refolding should be considered for all proteins in general. It would be advisable to remove such a tag for the study of protein thermostability.

Acknowledgments

This work was supported by a grant from National Sciences and Engineering Research Council of Canada (NSERC) awarded to M.B, no. 138654. D.M.C was supported by a fellowship from the National Sciences and Engineering Research Council of Canada (NSERC).

References

1. Terpe, K. Overview of tag protein fusions: From molecular and biochemical fundamentals to commercial systems. *Appl. Microbio. Biotechnol.*, 2003, 60, 523-533.
2. Hedhammar, M.; Graslunf, T.; Uhlén, M.; Hober, S. Negatively charged purification tags for selective anion-exchange recovery. *Prot. Eng. Des. Sel.*, 2004, 17, 779-786.
3. Carson, M.; Johnson, D.H.; McDonald, H.; Brouillette, c., Lawrence, J.O. His-tag impact on structure. *Acta Cryst D.*, 2007, 063, 295-301.
4. Derewenda, Z.S. The use of recombinant methods and molecular engineering in protein crystallization. *Methods.*, 2004, 34, 354-363.
5. Woestenenk, E.A.; Hammarstrom, M.; van den Berg, S.; Hard, T.; Berglund, H. His-tag effect on solubility of human proteins produced in *Escherichia coli*: a comparison between fluor expression vectors. *J. Struct. Funct. Genomics.*, 2004, 5, 217-229.
6. Amor-Mahjoub, M.; Suppini, J.P.; Gomez-Vrielyunck, N.; Ladjimi, M. The effect of the hexahistidine-tag in the oligomerization of HSC70 construct. *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sei.*, 2006, 844, 328-334.
7. Lee, Y-N.; Su, M-S.; Huang, T-H.; Shaw, J-F. C-terminal Histagging results in substate specificity changes of the thioesterase 1 from *Escherichia coli*. *J. Am. Oil. Chem. Soc.*, 1999, 76, 1113-1118.
8. Sayari, A.; Mosbah, H.; Verger, R.; Gargouri, Y. The N-terminal His-tag affects the enantioselectivity of staphylococcal lipases: A monolayer study. *J. Col./Int. Sei.*, 2007, 3/3, 261-267.
9. Consalvi, V.; Chiaraluce, R.; Giangacomo, L.; Scandurra, R.; Christova, P.; Karshikoff, A., Knapp,S., and Ladenstein,R. Thermal unfolding and conformational stability of the recombinant domain II of glutamate dehydrogenase From the hyperthermophile *Thermotoga maritima*. *Prot. Eng.*, 2000, 13(7), 501-507.
10. JunGang, L.; KeGui, Z.; and Wen Jiin, Han. Cloning and biochemical characterization of a novel lipolytic gene from activated sludge metagenome, and its gene product. *Microb. Cell Fact.*, 2010, 9, 83-87.

11. Bertolini, M.C.; Schrag, J.O.; Cygler, M.; Ziomek, E.; Thomas, D.Y.; Vernet, T. Expression and characterization of *Geotrichum candidum* lipase 1 gene. *Eur. J. Biochem.*, 1995, 228, 863-869.
12. Blank, K.; Morfill, J.; Gump, H.; Gallb, H.E. Functional expression of *Candida antarctica* lipase B in *Escherichia coli*. *J. Biotech.*, 2006, 25, 474-483.
13. Charbonneau, D.M.; Meddeb-Mouelhi, F.; Beauregard, M. A novel thermostable carboxylesterase [from *Geobacillus thermodenitrificans*: Evidence for a new carboxylesterase family. *J. Biochem.*, 2010, 148, 299-308.
14. Smith, P.K.; Krohn, R.I.; Hermanson, G.T.; Mailla, A.K.; Gartner, F.H.; Provenzano, M. O.; Fujimoto, E. K.; Goeke, N. M.; Olson, J. J.; and Klenk, D. C. Measurement of protein using bicinchoninic acid. *Anal. Biochem.*, 1985, 150, 76-85.
15. Sreerama, N.; and Woody, R.W. Estimation of the number of α -helical and β -strand segments in proteins using CD spectroscopy *Protein Sci.*, 1999, 8, 370-380.
16. Sreerama, N.; Manning, M.C.; Powers, M.E.; Goldenberg, D.M.; and Woody, R.W. Tyrosine, phenylalanine and disulfide contribution to the circular dichroism of proteins: circular dichroism spectra of wild-type and mutant bovine pancreatic trypsin inhibitor. *Biochemistry.*, 1999, 33, 10814-10822.
17. Sreerama, N.; and Woody, R.W. Estimation of protein secondary structure [from circular dichroism spectra: Comparison on CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.*, 2000, 287, 252-260.
18. Pace, C.N.; Shirley, B.A.; Thomson, J.A. In: Creighton, T.E., Ed. Protein Structure- A practical approach, IRL Press: Oxford, 1990; 311-329.
19. Wear, M.A.; Patterson, A.; and Walkinshaw, M.D. A kinetically trapped intermediate of FK506 binding protein forms *in vitro*: chaperone machinery dominates protein folding *in vivo*. *Prat. Exp. Purif.*, 2006, 51, 80-95.

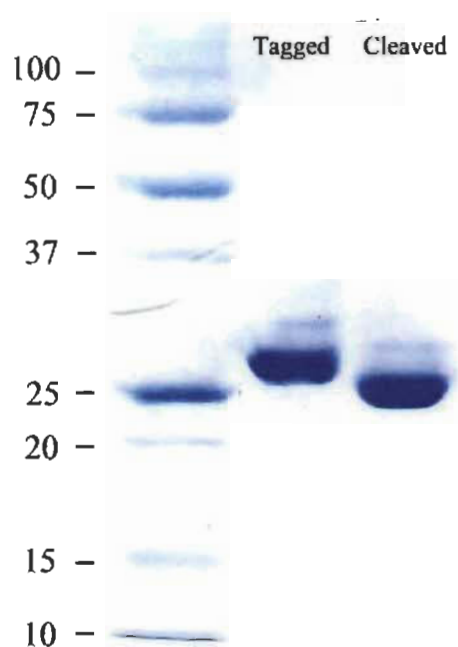


Figure 1. SDS-PAGE for the tagged and cleaved EstGtA2. First lane shows a molecular weight marker (precision plus), and lanes 2 and 3 show the tagged (~ 29.5 kDa) and cleaved protein (~ 27.4 kDa) respectively.

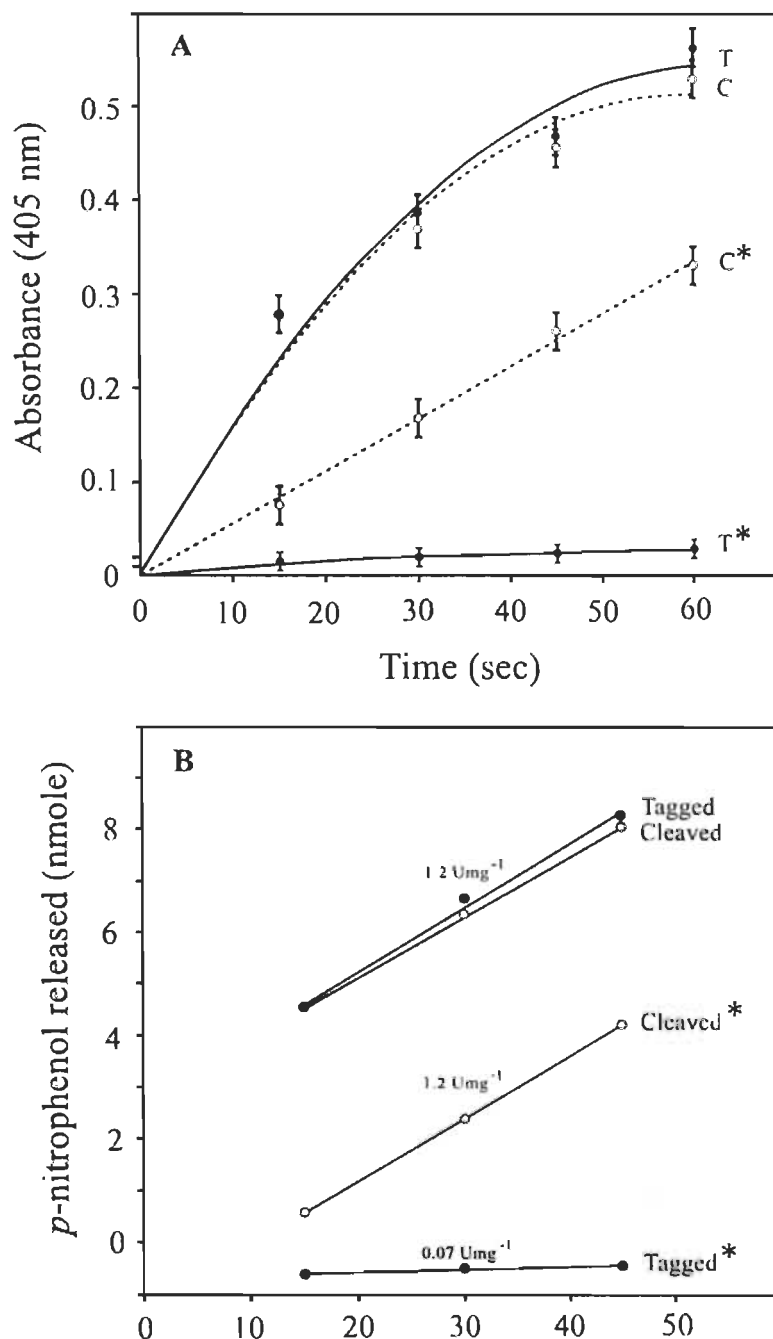


Figure 2. Enzyme activity and stability for hydrolysis of *p*-nitrophenyl caprylate. A. Absorbance at 405 nm measured at 25°C as function of time for EstGtA2 cleaved (open circles) and tagged (filled circles). B. Amount of *p*NP released, derived from the absorbance at 405 nm. A total of 5 μ g of protein was assessed with 10 nmole of substrate in 20 mM sodium phosphate pH 8 for 60 sec. Specific activity was calculated as described in Materials and methods. Asterisks indicate enzyme activity (cleaved and tagged) at 25°C after 15 min incubation at 80°C.

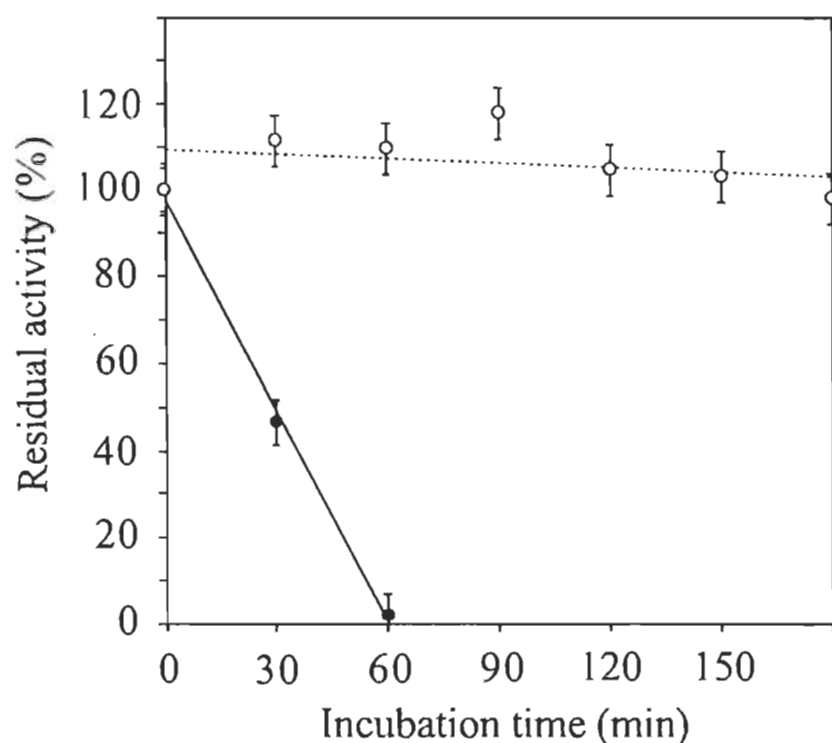


Figure 3. Impact of the N-terminal purification tag peptide on thermal stability of EstGtA2. Activity measured after various period of incubation at 80°C. Residual activity measured at 25°C (10 min reaction) for EstGtA2 cleaved (open circles) and tagged (filled circles) after incubation of protein samples. Enzyme (10 μ g) was assayed with 50 μ M p-nitrophenyl caprylate in 20 mM sodium phosphate pH 8. The amount of pNP released and residual activity were calculated from the absorbance at 405 nm.

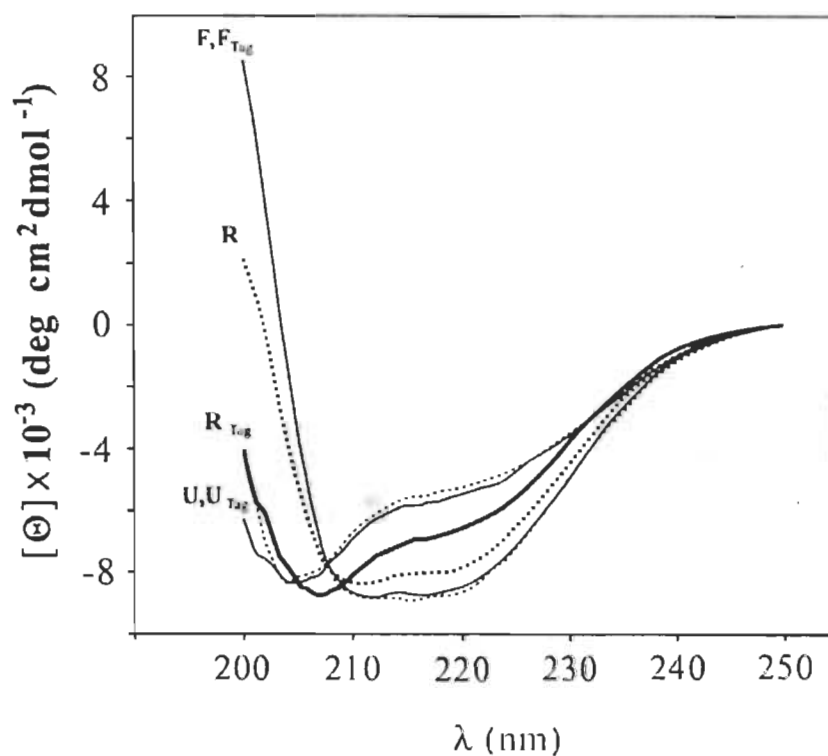


Figure 4. Effect of the N-terminal tag peptide on conformational reversibility of EstGtA2 after thermal unfolding. Molar ellipticity was recorded from 200-250 nm with 0.42 mg/mL of protein in 20 mM sodium phosphate, pH 8. Dashed lines show CD spectra for the cleaved protein and solid lines show spectra for the tagged protein. F and F_{Tag} indicate spectra recorded at 30°C, U and U_{Tag} are spectra measured at 80°C and R, R_{Tag} are CD spectra measured after cooling from 80 to 30°C.

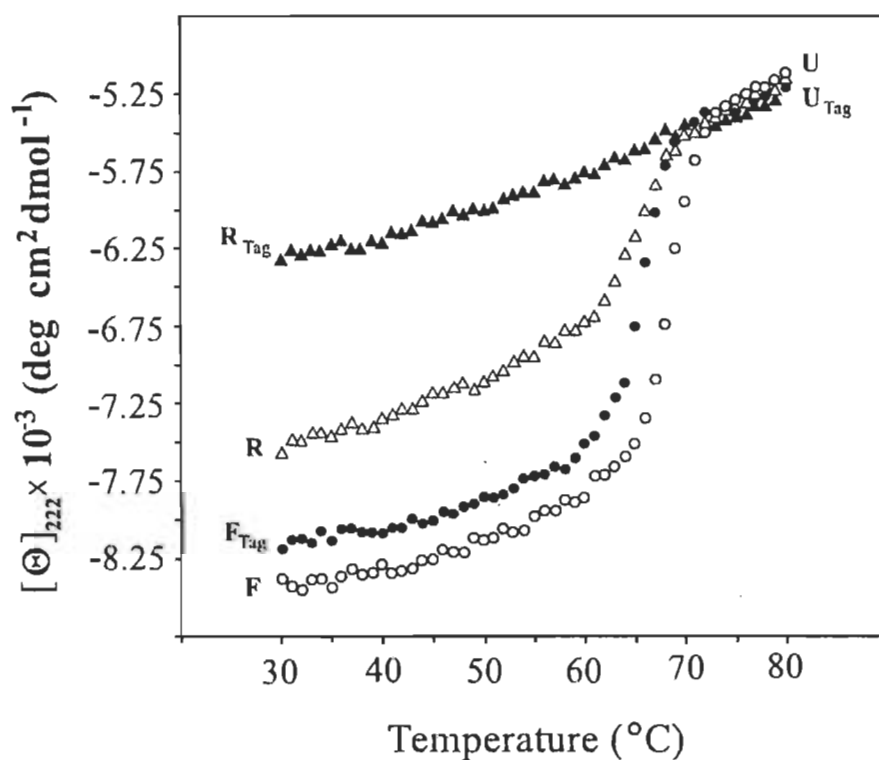


Figure 5. Impact of the N-terminal tag peptide on refolding of EstGtA2. Thermal unfolding and refolding profiles for the tagged and cleaved proteins. Molar ellipticity was recorded at 222 nm with 0.42 mg/mL of protein in 20 mM sodium phosphate, pH 8 from 30 to 80°C (F → U) and cooled from 80 to 30°C (U → R). Open circles and triangles and close circles and triangles show the EstGtA2_{cleaved} and EstGtA2_{tagged} profiles respectively.

CHAPITRE V

ROLE OF KEY SALT BRIDGES IN THERMOSTABILITY OF *G. THERMODENITRIFICANS* ESTGTA2: DISTINCTIVE PATTERNS WITHIN THE NEW BACTERIAL LIPOLYTIC ENZYME FAMILY XV

Ce chapitre est présenté sous forme de manuscrit et a fait l'objet d'une publication dans le journal scientifique PLoS One « Public Library of Science ».

David M. Charbonneau and Marc Beauregard. 2013. Role of key salt bridges in thermostability of *G. thermodenitrificans* EstGtA2: Distinctive patterns within the new bacterial lipolytic enzyme family XV. PLoS One (submitted).

5.1 Contribution des auteurs

La conception des expériences a été réalisée par David Charbonneau. Les différentes expérimentations ainsi que l'analyse des résultats ont été effectuées par David Charbonneau en collaboration avec Marc Beauregard. David Charbonneau a écrit en entier le manuscrit qui a été révisé par Marc Beauregard avant sa soumission finale dans le journal PLoS One.

5.2 Résumé de l'article

Les enzymes lipolytiques bactériennes ont initialement été classées en huit familles distinctes définies par Arpigny et Jaeger (famille I-VIII). Récemment, la découverte de nouvelles enzymes lipolytiques a permis d'étendre cette classification à quatorze familles (familles I-XIV). Nous avons déjà suggéré que la nouvelle EstGtA2 de *G. thermodenitrificans* appartenait à un nouveau groupe d'enzyme lipolytique que nous avons appelé N'. Nous proposons que ce groupe, qui diverge toujours des quatorze familles identifiées, appartienne à une quinzième famille d'enzymes lipolytiques

bactériennes (famille XV) et proposons des critères pour reconnaître les enzymes appartenant à cette nouvelle famille. Nous proposons aussi que le groupe N', originalement identifié forme en fait une sous-famille à l'intérieur de la famille XV. Le groupe N' montre une série de ponts salins strictement conservés, lesquels semblent jouer un rôle important dans la stabilité des enzymes de la sous-famille N'. Nous avons entrepris d'étudier le rôle de ces ponts salins dans la thermostabilité de EstGtA2 en utilisant une approche de mutagenèse combinatoire de type « alanine-scanning ». Les résidus basiques impliqués dans ces ponts salins (R ou K) ont été remplacés par des résidus alanine de façon combinatoire. 5 simples, 3 doubles, 3 triples et 3 quadruples mutants ont été produits et caractérisés. Les résultats montrent que EstGtA2 tolère plusieurs de ces mutations non conservatives pour adopter un repliement proche de celui du type sauvage, par contre, quoi que l'enzyme se replie à basse température, l'accumulation de plus de deux de ces mutations se traduit par une inactivation presque totale à haute température et à la température optimale du type sauvage. Nous avons aussi identifié une position particulière, où un pont salin est observé dans la famille XV et XIII, entre deux résidus chargés coévoluant en position ($i - 2$ et $i - 4$) de l'aspartate et de l'histidine catalytique. Quoique conservée, l'identité des résidus formant ce pont et la polarité de l'interaction électrostatique est fréquemment renversée au cours de l'évolution. Nous avons utilisé cette observation pour introduire une mutation rationnelle qui a permis d'améliorer la thermostabilité de EstGtA2 à pH alcalin. Nos résultats montrent que les ponts salins ciblés constituent des motifs de coévolution caractéristique pour l'assignation génomique et des cibles intéressantes pour l'optimisation d'enzymes de la famille XV et XIII.

5.3 Quatrième article scientifique

Role of key salt bridges in thermostability of *G. thermodenitrificans* EstGtA2: Distinctive patterns within the new bacterial lipolytic enzyme family XV

David M. Charbonneau^{1,2} and Marc Beauregard^{1,2*}

1. Département de chimie-physique, Centre de recherche sur les matériaux lignocellulosiques, Université du Québec à Trois-Rivières, 3351 Boul. Des Forges, C.P. 500 Trois-Rivières (Québec) G9A 5H7, Canada
2. PROTEO, Université Laval, 2705 Boul. Laurier, Ste-Foy (Québec) G1V 4G2, Canada

Running title:

Role of conserved salt bridges in carboxylesterase EstGtA2

* To whom correspondence should be addressed:

Prof. Marc Beauregard

Centre de recherche sur les matériaux lignocellulosiques, Université du Québec à Trois-Rivières, C.P. 500, Trois-Rivières (Québec), G9A 5H7.

Tel: (819) 376-5011 # 3354.

Fax: (819) 376-5084.

E-mail: marc.beauregard@uqtr.ca

PLoS One. 2013 Oct 8;8(10):e76675. doi: 10.1371/journal.pone.0076675.

Article available online at: <http://www.ncbi.nlm.nih.gov/pubmed/24116134>

Summary

Bacterial lipolytic enzymes were originally classified into eight different families defined by Arpigny and Jaeger (families I-VIII). Recently, the discovery of new lipolytic enzymes allowed for extending the original classification to fourteen families (I-XIV). We previously reported that *G. thermodenitrificans* EstGtA2 (access no. AEN92268) belonged to a novel group of bacterial lipolytic enzymes. Here we propose a 15th family (family XV) and suggest criteria for the assignation of protein sequences to the N' subfamily. Five selected salt bridges, hallmarks of the N' subfamily (E3/R54, E12/R37, E66/R140, D124/K178 and D205/R220) were disrupted in EstGtA2 using a combinatorial alanine-scanning approach. A set of 14 (R/K→A) mutants was produced, including five single, three double, three triple and three quadruple mutants. Despite a high tolerance to non-conservative mutations for folding, all the alanine substitutions were destabilizing (decreasing T_m by 5 to 14°C). A particular combination of four substitutions exceeded this tolerance and prevents the correct folding of EstGtA2, leading to enzyme inactivation. Although other mutants remain active at low temperatures, the accumulation of more than two mutations had a dramatic impact on EstGtA2 activity at high temperatures suggesting an important role of these conserved salt bridge-forming residues in thermostability of lipolytic enzymes from the N' subfamily. We also identified a particular interloop salt bridge in EstGtA2 (D194/H222), located at position $i - 2$ and $i - 4$ residues from the catalytic Asp and His respectively which is conserved in other related bacterial lipolytic enzymes (families IV and XIII) with high tolerance to mutations and charge reversal. We investigated the role of residue identity at position 222 in controlling stability-pH dependence in EstGtA2. The introduction of a His to Arg mutation led to increase thermostability under alkaline pH. Our results suggest primary targets for optimization of EstGtA2 for specific biotechnological purposes.

Introduction

The understanding of sequence-encoded information has never been as important as in this post-genomic era where protein purification and characterization is dramatically outpaced by the volume of DNA sequences deposited in various databanks. Our inability to predict protein behaviour using sequence information drastically slows the pace of discovery. By improving our understanding of sequence features that control enzymatic properties or permit enzyme assignment to a characterised group of related enzymes, we will achieve more accurate extraction of protein character information from genomic data.

Carboxylesterases (EC 3.1.1.1) and lipases (EC 3.1.1.3) are lipolytic enzymes belonging to the class of hydrolases which catalyze the hydrolysis or synthesis of ester bonds in lipids. They adopt the α/β hydrolase fold and bear a catalytic triad formed by Ser-Asp/Glu-His, but differ in term of substrate specificity [1, 2]. Carboxylesterases catalyze the hydrolysis of ester bonds in short-chain and partially water-soluble substrates while lipases prefer long-chain triglycerides. They are widely distributed in nature from prokaryotes to mammals and involved in many physiological roles including carbon utilization, detoxification and processing of neurotransmitters to name but a few examples. Prokaryotic lipolytic enzymes have been assigned to eight different families (I-VIII) based on sequence identity and biochemical properties by Arpigny and Jaeger [3]. Triacylglycerol lipases are grouped in family I, while carboxylesterases are grouped in families II-VIII. Recently, the discovery of new bacterial lipolytic enzymes led to the establishment of new families that diverge from the original 8-family classification (families IX-XIV), [4]. The fact that Est30 (CE_{GK}) from *Geobacillus kaustophilus* (99% identity with Est30 from *G. stearothermophilus*, PDB code. 1TQH) did not fit in the 8-family classification was reported [5, 6]. It was argued that CE_{GK} was different from families IV and VI on the basis of not only the low sequence identity, but also on the lid/cap structures and the number of strands in the core β -sheet. The Est30-like enzymes form the family XIII. We previously suggested that the novel *Geobacillus thermodenitrificans* EstGtA2 diverged from the classification proposed by Arpigny and Jaeger, and was instead a representative of a new family of thermostable lipolytic

enzymes named N' [7]. Despite important similarities between N' and Est30-like enzymes (family XIII), the N'-related enzymes form a clearly distinct group including monoacylglycerol lipases. A similar conclusion was reached for LipS (from *Symbiobacterium thermophilum*), [8]. Interestingly, an earlier, inclusive classification by Fischer allows assigning Est30-like enzymes (family XIII) to family abH11.01 (Fischer *et al.*, 2003). Similarly, the N' group can be assigned to a cluster included in family abH11.03. The recently proposed LipS group would also belong to abH11.03 family as proposed by Fischer *et al.* [9]. Here we will revisit the various classifications or family names suggested for these lipolytic enzymes and attempt to reconcile them by proposing their assignment to the new 15th family (family XV), named chronologically based on the extension of the Arpigny and Jaeger's classification.

Thermal stability of lipolytic enzymes has become a practical concern as a result of their applications in different biotechnological processes [10]. Due to their high thermal stability, carboxylesterases and lipases from thermophiles have been and are still intensively studied. Considerable efforts have been made to understand mechanisms involved in determining protein thermostability. Electrostatic interactions between opposite charge residues have gained attention as important determinants of thermostability in thermophilic proteins, which often exhibit more salt bridges or charged networks than their mesophilic counterparts [11-13]. Multiple salt bridges are accepted as one of the most important features that contribute to enzyme thermostability. A statistical analysis of 18 non-redundant families of thermophilic and mesophilic proteins revealed similar hydrophobicities, compactness, polar and non-polar contributions to surface areas, main-chain and side-chain hydrogen bonds. At variance with such conservation of features, the number of salt bridges and side-chain-side-chain hydrogen bonds increased in the majority of thermophilic proteins. In addition, comparison of thermophilic and mesophilic homologous proteins indicated that arginine and tyrosine are significantly more frequent in thermophilic versions [14].

Although many studies have suggested a potential impact of surface electrostatic interactions on protein thermostability, the precise contribution of surface salt bridges to

protein stability has long been disputed and is still debated [15-18]. Since formation of a salt bridge depends on the pH of the environment and on the pK_a of the charged groups involved in the ion pair ($-\text{NH}_3^+ \cdots ^- \text{OOC}-$), the contribution of a particular surface salt bridge to thermal stability may vary according to pH conditions. Therefore, depending of the identity of the charged residues involved, the contribution for thermal stability becomes pH-dependent. Other factors are also at play in determining the extent of stability afforded by a bridge [18-20].

Historically, stability determinants have been largely inferred from comparisons of homologous proteins from distant mesophilic vs. thermophilic organisms. Unfortunately, most attempts to use such putative stability determinants in a different context have yielded unpredictable results [21, 22]. The overall impact of the introduction of new salt bridges on enzyme thermotolerance can also be negative, depending on the contribution of short- and long-range interactions and other factors [23-26]. Parameters such as level of exposure to solvent, relative orientation of the side-chains, short- and long-range interactions play a determinant role, and may vary widely when comparing distant proteins.

The ability to predict the impact of new salt bridges as stabilization devices in mutant proteins using canonical principles derived from distant proteins has remained elusive. But comparison of closely related proteins with differences in thermostability or other properties should reveal structure-properties relationships of direct relevance for the proteins chosen [27]. In this respect, the high level of identity between enzymes included in the N' cluster (85-100% identity) makes them ideal subjects for the investigation of conserved structural determinants for enzyme properties, including thermostability.

Here we identified eight salt bridges in EstGtA2 that are highly conserved in the N' cluster. Of these, five salt bridges involving Arg or Lys residues are strictly conserved and might be useful in as potential sequence indicator for the assignation of newly discovered enzymes to the N' subfamily. Our results show that these five selected

basic residues are involved in controlling thermostability in EstGtA2. We also explored the dependence of stability on pH by mutating a particular interloop salt bridge found in proximity to the catalytic site. Our work on EstGtA2 opens vista on structural determinants for thermostability of lipolytic enzymes from the N' subfamily.

Materials and Methods

Strains and plasmids

The EstGtA2 ORF (GenBank accession no. JN031579) was amplified by PCR from *G. thermodenitrificans* strain CMB-A2 (GenBank accession no. GQ293454) [7], and cloned into BL21-Gold(DE3)pLysS strain: *E.coli* B F⁻ ompT hsdS(r_B⁻ m_B⁻) dcm⁺ Tet^r gal λ(DE3) endA Hte [pLysS Camr] (Stratagen) using the pET28(a+) expression vector (Novagen).

Site-directed mutagenesis

Mutations were introduced in EstGtA2 using the QuickChangeII site-directed mutagenesis kit (Stratagene). PCR were carried out with the *Pfu* ultra HF DNA polymerase (Stratagene). The mutated PCR products were treated with *DpnI* and cloned into *E. coli* BL21 using heat shock transformation. Transformed cells were selected on LB-agar containing 50 µg/mL kanamycin. Selected clones were picked and grown in 5 mL LB 50 µg/mL kanamycin overnight at 37°C with agitation (200 rpm). The plasmids were extracted and purified from 3 mL cultures using the Miniprep kit (Qiagen) and 1 mL of each culture was stocked in 20% glycerol and stored at -80°C. Sequencing was carried out using the T7 terminator primer at the Biomolecular platform, Université Laval (Québec). Specified mutated plasmids were used as template at each step for combination of mutations, and additional site-directed mutagenesis. The primers used for cloning and directed mutagenesis are listed in Table S1 in File S1.

Protein expression and purification

Transformed cells with genes coding for wild type and mutant proteins were grown overnight in 5 mL LB containing 50 µg/mL kanamycin for 16 h at 37°C with agitation at 175 rpm. Volumes of 125 mL LB (50 µg/ mL kanamycin) were inoculated with 0.125 mL of the respective pre-cultures and were grown at 37°C until $OD_{600} \approx 0.6$ and then induced with 0.5 mM IPTG (final concentration) for 16 h at 20°C with agitation (150 rpm). Cells were harvested by centrifugation and stored at -80°C. Cells were then thawed and resuspended in 15 mL of lysis buffer (50 mM sodium phosphate, 10 mM imidazole, 150 mM NaCl, pH 7.5 and 1 mg/mL lysozyme was added) and lysed by sonication for 30 sec at 50 pulses/sec. The lysate was clarified by centrifugation for 30 min at 15 krpm and the expressed His₆-tagged recombinant proteins were purified from the soluble fraction under native conditions using IMAC (Ni-NTA beads, Qiagen). The nickel resin was washed with 30 mL of buffer. Bound His₆-tagged proteins were eluted from the nickel resin with 5 mL of elution buffer (20 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole, pH 7.5). Proteins were dialyzed 48h hours at 4°C against 2L of sodium phosphate buffer (20 mM), pH 8 using 3,500 kDa MWCO dialysis membranes. The protein concentration was determined by the bicinchoninic acid (BCA) method. The purity of each preparation was verified by SDS-PAGE.

Enzyme activity measurements

Enzyme activity was assessed in a final volume of 0.2 mL in 96-wells microplates using a Synergy Mx, Biotech microplate reader. Each reaction contained a final concentration of 10 µg/mL (1 µg) of purified protein, with 100 µM (20 nmole) *p*-nitrophenyl octanoate in 20 mM buffer. The buffers used were sodium citrate/citric acid (pH 3-5), sodium phosphate (pH 6-8), Tris-HCl pH9 and CAPS/NaOH (pH 10-12). Specific activity was determined by measuring the amount of para-nitrophenol (*p*NP) released after hydrolysis of *p*NP-octanoate at temperatures ranging from 25 to 65°C. The amount of *p*NP produced was recorded by monitoring the absorbance at 405 nm (pH ≥ 7). The absorbance at 340 nm was also monitored under acidic conditions.

Standard curves for *p*-nitrophenol absorbance (at each pH and temperature) as function of concentration allowed determining the initial rate of hydrolysis and enzyme specific activity at each pH and temperature. One unit of enzyme activity was defined as the amount of *p*NP (μ mole) released by minute per milligram of pure protein. Where specified (H222R and wild type) enzymes were incubated at various temperatures for 10 minutes, then cooled down to 25°C for measurements.

CD spectroscopy

Far-UV (178-250 nm) and near-UV (250-320 nm) CD spectra were recorded in a Jasco J-720 spectropolarimeter under nitrogen atmosphere with protein concentration of 0.5 mg/mL in 20 mM sodium phosphate pH 8 or CAPS pH 10 using a 0.01 and 1 cm quartz cell respectively. The data were converted into molar ellipticity taking into account the path length and the molar concentration of residues (0.0045 M). For calculation of secondary structures, the far-UV CD data were converted into $\Delta\epsilon$. The protein secondary structures were calculated using CDSSTR (spectra set no. 3 from 190-260 nm) [28-29]. Thermal unfolding/refolding transitions were measured with a 0.1 cm path length jacketed quartz cell connected to a Neslab RTE-111 circulating water bath containing 50% ethylene glycol. The transitions were monitored by the variation of CD signal at 222 nm between 25 to 90°C upon heating and cooling at a rate of 0.8°C/min. Transitions were evaluated using a nonlinear least squares fit assuming a two-state model with sloping pre- and post-transitional base lines. The standard errors for T_m calculated from the analysis of the individual melting profiles were smaller than 0.1°C. Extrapolation and least-square analysis from the pre- and post-transition regions allowed determination of the equations for y_F and y_U . Thermal unfolding curves were fitted on a two-state model ($F \leftrightarrow U$). The equilibrium constant was determined at each temperature and the thermodynamic parameters were determined according to the following equations, [30]:

$$\text{eq1. } f_U = (y_F - y)/(y_F - y_U)$$

$$\text{eq2. } K = f_U/(1 - f_U) = f_U/f_F = (y_F - y)/(y - y_U)$$

The free energy change associated with thermal unfolding was calculated at each temperature by the following equation:

$$\text{eq3. } \Delta G = -RT \ln K = -RT \ln[(y_F - y)/(y - y_U)].$$

The apparent T_m values were then determined by plotting ΔG as function of temperature where $T_m = T$ at $\Delta G = 0$, the midpoint of thermal unfolding curves, where $\Delta G(T_m) = 0 = \Delta H_m - T_m \Delta S_m$. This relation gives parameters to calculate the free energy change at any temperature $\Delta G(T)$, the equation can be given by:

$$\text{eq4. } \Delta G(T) = \Delta H_m (1 - T/T_m) - \Delta C_p [(T_m - T) + T \ln(T/T_m)]$$

Fluorescence spectroscopy

Protein intrinsic fluorescence was measured using a Carry Eclipse spectrofluorimeter with a protein concentration of 0.05 mg/mL in 20 mM phosphate buffer pH 8. The fluorescence emission spectra were recorded from 300 to 600 nm after excitation at 287 nm. For chemical denaturation experiments, 1 mL GuHCl (from 1 to 6 M) was mixed with 1 mL of protein (0.1 mg/mL). Fluorescence was recorded at room temperature with constant protein concentration of 0.05 mg/mL and final GuHCl concentration of 0-3 M in 20 mM sodium phosphate pH 8 using a 1 cm quartz cell.

In silico analyses

DNA and translated amino acid sequences were edited with Clone Manager Professional Suite version 7.03. Similarity searches were performed with BLAST 2.0 program [31] and sequences for comparative studies were retrieved from the GenBank database [32] via NCBI Entrez at <http://www.ncbi.nlm.nih.gov/Entrez/>. Multiple sequence alignments were performed with ClustalW2 [33]. The alignment postscripts were generated using ESPript 2.2 [34]. Phylogenetic analyses were carried out with MEGA 5.1 [35] using the neighbour-joining method [36]. The EstGtA2 3D structure

was predicted by homology modelling using the ESyPred3D program [37] using *B. sp* H-257 MGL crystal structure (PDB no. 3RM3) as template (89% identity with EstGtA2). Image of the resulting 3D model was generated using PyMOL and the quality of the model was evaluated using WHAT_CHECK and PROCHECK [38] available in the SWISS-MODEL workspace [39]. Computation of continuum electrostatic and evaluation of salt bridges formation based on the EstGtA2 model was performed using the ESBRI interface [40]. The accessible surface area (ASA) was calculated using the EstGtA2 3D structure model (PDB file) using ASA-View and GetArea (University of Texas, Medical Branch) and the pK_a values of charged groups involved in putative salt bridges formation was evaluated using the PROPKA server [41].

Results

pH-dependent stability of EstGtA2

EstGtA2 unfolding and refolding was probed as function of pH by monitoring the CD signal at 222 nm as function of temperature. The melting temperatures (T_m) for the wild type enzyme ranged from 64 to 69°C between pH 5 and 9. The lowest melting temperatures were recorded for pH 4 and 10 (56 and 54.5°C respectively, pH limits allowing for folding), Figure 1. Melting temperature decreased slightly at pH 8, which corresponds to optimal pH for activity (at optimal temperature of 50°C). At 25°C, the optimal pH for activity is observed at pH 9-10 and the protein is still active at pH 11, however completely inactivated at pH 12. Partial reversibility of thermal unfolding was only observed under alkaline pHs (pH 7-10). The conformational stability of EstGtA2 calculated from the CD transitions ranged from 8-17 kcal/mol depending on pH. Under acidic condition, aggregation appeared at high temperature leading to irreversible thermal unfolding (data not shown). The high sensitivity to pH suggests that ionized residues and/or salt bridges play a role in EstGtA2 stability.

Structural model of EstGtA2

In order to identify charged residues likely to impact on stability, the tridimensional structure of EstGtA2 was modeled using the X-ray crystal structure of MGL H-257 (PDB code. 3RM3, 89% residue identity with EstGtA2) [42-44]. As shown in Figure 2, the α/β core comprises a β -sheet formed by 7 β -strands (numbered β 2- β 8) surrounded by 6 helices (numbered α 1- α 6). The catalytic triad for EstGtA2 and MGL H-257 is unambiguously identified as Ser97, Asp196 and His226 and the oxyanion hole is formed by Met98 and Phe29. The residues of the binding pocket identified in MGL H-257 are all conserved in EstGtA2 (Figure S1). In addition to the canonical α/β core, a 45 residues cap domain (I125-T164) is found above the main α/β core, a structural feature found in many α/β hydrolases. The cap region is predicted to harbour a small α -helix (A125-A130) and two short anti-parallel β -strands (T141-D143), and (L161-P163) connected by linker regions. A separate α -helix (62-67) interacts with the large cap insertion moiety (I125-T164) forming the cap structure (Figure 2). Comparison of the model with MGL H-257 suggests a high conservation of the residues in the cap region, the binding pocket and the β -sheet. Interestingly, the 27 amino acids that are different in EstGtA2 compared with MGL H-257 are concentrated in α -helices and loops of the core of EstGtA2 (Figure S2). These mutations may contribute to the flexibility of the binding pocket or the catalytic cleft, which in turn would explain the difference of specificity between EstGtA2 and MGL H-257 observed (EstGtA2 is able to hydrolyze tributyrin while MGL H-257 cannot) [43]. The Ramachandran plot and the model refinement statistics are presented in supporting information (Figure S3). The conservation of 89% of residues between EstGtA2 and MGL H-257, and the quality assessment of the model, supports the hypothesis that EstGtA2 and MGL H-257 adopt a very similar structure.

Multiple conserved salt bridges in EstGtA2

Conserved salt bridges were predicted in the structures of EstGtA2 and MGL H-257 using ESBRI with a cut-off distance of 4 Å. In addition, the crystal structure for MGL H-257 was solved from protein crystals obtained at pH 6.5-7.0, which is close to the pH value where these salt bridges were studied in EstGtA2 (pH 8). Eight salt bridges

were suggested in EstGtA2 and MGL H-257: E3-R54 (3.3 Å), E12-R37 (3.9 Å), E78-H110, E66-R140 (3.6 Å), D124-K178 (3.3 Å), D148-H197, D194-R/H222 (4.0 Å) and D205-R220 (3.6 Å). Seven out of those eight salt bridges are identical in EstGtA2 and MGL H-257. In the eighth, an arginine in MGL replaces the histidine found in EstGtA2 at position 222 (D194-H222 in EstGtA2; D194-R222 in MGL 257). This salt bridge links the loop between strand $\beta 7$ and helix $\alpha 5$ (displaying the catalytic Asp residue) to the loop between strand $\beta 8$ and helix $\alpha 6$ (harbouring the catalytic histidine, see Figures 2 and 3) and is conserved at positions $i - 2$ and $i - 4$ from the respective catalytic residues.

The calculated solvent accessibility for the salt bridges-forming charged residues predicts that most of these salt bridges are relatively buried except for the interloop salt bridge which is highly exposed at the protein surface (Table S2 in File S1). The E3-R54 salt bridge is the most buried ion pair; as such one might expect a more important contribution to stability from this electrostatic interaction. The theoretical prediction of the pK_a values for the residues involved in salt bridges range from 3.78 to 4.2 for glutamates and from 3.56 to 3.74 for aspartates. The pK_a values for arginine and lysine residues ranged from 11.09 to 13.85 while for histidines they varied between 6.03 and 6.64. It should be noted that the three salt bridges involving histidine in EstGtA2 are not considered for these predictions, since they are not expected to be formed in the pH range (pH 8-10) studied here.

Salt bridges from closely related enzymes were also examined. Salt bridge composition for EstGtA2 (N' subfamily), LipS and Est30 (N subfamily) is shown in Figure 4 and Table S3 in File S1. Three distinctive patterns are observed, wherein salt bridge-forming residues show different separations in sequence and in space. These different patterns are each conserved within closely-related homologs. Five bridges were exclusively conserved in lipolytic enzymes from the N' subfamily: E3-R54, E12-R37, E66-R140, D124-K178 and D205-R220, but absent from the recently characterised LipS (PDB code. 4FBL) or from the closely related N subfamily Est30 (PDB code. 1TQH) as shown in Figures 3 and 4. An additional multiple sequence alignment was built for

N subfamily representatives (Figure 3, panel B) and revealed a series of 6 bridges found exclusively in N subfamily. Such exclusive bridge composition appears to be a hallmark for quick recognition of N'- or N-related enzymes from genomic analysis.

Salt bridges as hallmarks of the N' subfamily

The first salt bridge among the five bridges conserved in the N' subfamily, E3-R54, links the N-terminal end of EstGtA2 to the loop after strand β 3. Three hydrogen bonds and electrostatic interactions between the guanidinium group of R54 and the carboxylate of E3 side-chain are predicted involving the resonance form and the positive charge delocalisation of the guanidinium group of R54 (Figure S4). Two additional hydrogen bonds are formed, one between the R54 guanidinium and the backbone carbonyl oxygen of H58 and one between the amino group of the R54 backbone and the OH group of the S76 side chain. In addition, the aliphatic moiety of R54 side-chain is buried in a hydrophobic cluster with two proline residues, (P13 and P33). The E12-R37 salt bridge also functions to link the N-terminal end of EstGtA2 to the core through a hydrogen bond found between the guanidinium group of R37 side-chain and the carboxylate of E12 side-chain.

Another noteworthy predicted salt bridge between E66-R140 links the large insertion of the cap domain to the separated helix of the cap formed by residues 61 to 67. In addition to salt bridge between the guanidinium of R140 and E66 carboxylate, residue R140 forms an additional hydrogen bond with the OH group of Y141. The salt bridge formed between D124 and K178 is the second most buried ion pair. A hydrogen bond is formed between the ϵ -amino group (NH_3^+) group of K178 and the carboxylate group of D124. Two additional hydrogen bonds are formed between D124 and N203, one between the backbone carbonyl oxygen of D124 and the NH_2 side-chain amide group of N203 and one between the backbone NH group of D124 and the oxygen of the side-chain amide group of N203. The aliphatic moiety of K178 side-chain also interacts with the V123 side-chain. The salt bridge D205-R220 contains only one hydrogen bond and electrostatic interaction between side-chains and links the strand β 8 to the helix α 5

(Figure S5). The five exclusive salt bridges displayed important interactions and might play a role in EstGtA2 stability.

Of the eight salt bridges found in EstGtA2 and MGL H-257, only two are conserved in the homolog LipS. The first one is D148-H197 (EstGtA2), corresponding to D179-H228 (in LipS), which links the cap domain to the α/β core. The second conserved salt bridge is located at the interloop position. Despite the low conservation of most salt bridges in EstGtA2 compared to enzymes from the closely related N subfamily (family XIII) or recently identified LipS [8], a particular interloop salt bridge appears to be conserved across most members of these recently identified families, but with high tolerance to mutations. A phylogenetic tree (Figure 5) shows the evolution and polarity reversal at this conserved salt bridge position within N and N' subfamilies, and for other distant bacterial α/β hydrolases. This salt bridge is formed between R225 and E253 in LipS, R191 and E219 in Est30, D194 and R222 in MGL 257 and D194 and H222 in EstGtA2 (Figures 3 and 4), to name a few. The polarity of the bridge is switched from acid-base polarity in the N' subfamily, to base-acid polarity in N subfamily (Figure 5). This salt bridge tethers loops directly connected to active site segments, and is conserved at residue positions $i-2$ and $i-4$ from the catalytic Asp and His respectively. The bridge is not perfectly conserved however (see for example YP_003253705 and YP_002885227 in Figure 5), suggesting that its presence is not absolutely necessary. This bridge was studied here considering its proximity to active site residues, and possible impact on catalytic properties. Given that histidine is the positive partner at this position in EstGtA2, we investigated a possible impact of replacing His by Arg on pH-dependent properties.

Combinatorial alanine-shaving of salt bridge-forming residues

Here the contribution to stability of the five salt bridges identified above (exclusively conserved in N' family) was studied further using combinatorial alanine-scanning mutagenesis. To this end, the bridges containing Arg or Lys residues as the positively charged partner (salt bridges E3-R54, E12-R37, E66-R140, D124-K178 and

D205-R220), were progressively disrupted by substituting alanine for K or R, Figure 2, Table 1. Single, double, triple and quadruple mutants were produced as depicted in Table 1. All mutants were active at room temperature (Figure 6) and correctly folded as suggested by near- and far-UV CD, except for the quadruple mutant M4c spectra (Figures 7 and 8). These mutations resulted in complete inactivation of the enzyme. All single mutations caused an increase in the activity at room temperature and moderate temperature (below 50°C, the optimum temperature for the wild type), perhaps due to an increased flexibility of the core compared to EstGtA2 at this temperature (see Figure 6 and Table 1). The increased activity observed for single mutants was also observed at 60°C and on *p*NP ester of longer acyl chain-length (*p*NP-C12). This appears to be due to a general improvement of activity rather than a shift in chain-length specificity (data not shown). The mutation R37A is the most activating one with a 1.2 and 1.5 fold increase activity at 25 and 60°C. At 60°C, the R54A mutation is the most destabilizing single mutation, as mentioned above, this bridge is relatively buried and involved in a network of interactions (Figure S4). Triple mutants were nearly as active as the wild type enzyme at 25°C, but were inactivated at 60°C where the wild type EstGtA2 retained more than 70% activity. Among the three quadruple mutants analysed, M4b was as active as EstGtA2 at 25°C and at variance with triple mutants, it remains active at 60°C (40% of wild type). M4a lost 70% of activity at 25°C, and became inactive at 60°C. The particular set of mutations leading to M4c led to total inactivation of this mutant (Table 1). In general, the R37 mutation appears to stimulate (or prevent inactivation) activity (in M1a, M4b), while R54 is present in less active mutants (in M1b, M3a, M3b and M4c).

Thermal denaturation curves were measured by monitoring the molar ellipticity at 222 nm as function of temperature and are expressed as fraction folded (Figure 9). The melting temperatures (T_m) were calculated by plotting ΔG as function of temperature, where $T_m = T$ at $\Delta G = 0$ (Figure S6). The thermodynamic parameters and activity measured for the wild type and mutants are shown in Table 1.

Despite the fact that all mutants (except M4c) folded correctly and retained activity at low temperature, all the mutations were significantly destabilizing for EstGtA2. As shown in Figure 10 and Table 1, the R37A, R54A, R140A, K178A and R220A single mutants were characterised by lower T_m (with ΔT_m ranging from - 5.6 to - 8.9°C). The double mutants had a decrease in T_m ranging from - 8.6 to - 13.6°C. Triple mutants were even less stable than the single and double mutants, with drops in T_m ranging from - 10.3 to - 13.8°C. The worst scenarios were observed with quadruple mutants, where decrease in T_m ranged from - 11 to - 14.3°C. Only single mutants retained activity above 60°C (see Figure 6). Interestingly, the triple mutant R140A/K178A/R220A (M3c) showed a reduced T_m of - 13.8°C but folded correctly and was active at room temperature (same observation for the two other triple mutants). However, they were much less stable and were inactivated at 60°C. Interestingly, when the R54A mutation was added to the three mutations of M3c to form the quadruple mutant R54A/R140A/K178A/R220A (M4c), the folding of the enzyme was compromised. The quadruple mutant M4a shows considerable conformational changes and was almost inactivated at room temperature, being completely inactivated at 50°C. The quadruple mutant M4b was much more active than M4a (almost equal to the wild type at 25°C) but was inactivated over 60°C (Figure 6). Interestingly the mutant M4b contains the mutation R37A which may play a compensatory role, while M4a contains R54A which seems to be the most destabilizing single mutation. When R54A is combined with R140A/K178A/R220A, protein folding is prevented. The double mutant M2c showed a reduction in T_m that is comparable to triple mutant M3c as well as M4b. It appears that mutation R140A which decreased T_m by 7.6°C in the single mutant M1c did not destabilize further the M3c or M4b derivatives, or that some compensatory mechanisms are at play in multiple mutants.

Misfolding of M4c

As shown in Figures 7 and 8, important modifications in the far- and near-UV CD spectra were observed for the quadruple mutant (M4c) compared to wild type and the other mutants. The molar ellipticity in the far-UV region shifted from the typical spectral

shape observed for α/β structures to a spectrum having the hallmarks of protein denaturation (negative peak at 203 nm, loss of signal in the range of 210-230 nm). Calculation of secondary structures for the native wild type enzyme predicted 35% α -helices, 20% β -sheets, 20% turns and 25% unordered structures. These values are close to those calculated with the atomic model of MGL 257 3RM3 (35% helices, 16% strands). The quadruple mutant calculations revealed a major reduction of α -helices (-23%) with a concomitant gain in beta and unordered structures (+7%) compared to wild type. The near-UV CD spectra that result from compactness in the surrounding of aromatic residues changed slightly for M3a, M3b, M3c, M4a, and M4b. This suggests some modification in the packing, a possible consequence of the change in bulk at the mutated position (R being much larger than A). This effect seemed to be more drastic for M4c which showed a major decreased in CD signal intensity at 290 nm (possibly corresponding to tryptophan). This observation suggests an increased mobility of the tryptophan side-chains possibly as a result of the unfolded state of this mutant.

The fluorescence spectra (Figure 10) also suggest misfolding of M4c as observed by CD. An observed shift of the fluorescence emission peak (λ_{\max}) from 320 to 355 nm indicates that hydration occurred in the hydrophobic core of the protein and/or that polar contacts arose in the immediate environment of the originally highly buried tryptophan residues Trp73 (0) and Trp83 (0.19) (where numbers in parentheses indicate solvent accessibility calculated in the folded structure). Chemical denaturation experiments were performed as shown in Figure 10, panel B. The wild type EstGtA2 shows a $[\text{GuHCl}]_{1/2} = 2.2$ M in comparison with 1.75M for M3c. The quadruple mutant M4c did not show any transition or shift in λ_{\max} indicating that tryptophan residues are already exposed to solvent in the buffer (in the absence of denaturant). Such results confirm that M4c is unfolded and consequently inactive.

Activation and specificity modification by R37A

Although the R37A mutation is destabilizing (lowering the T_m), it increased EstGtA2 activity at low temperature and shifted the optimal temperature from 50 to

55°C. The activation effect seems to be coupled with a change in specificity. The wild type EstGtA2 hydrolyzes tributyrin (C4) but cannot hydrolyze long-chain triglycerides (TAG) such as olive oil (C16-18). Surprisingly, the mutants M1a (R37A) showed activity on emulsified olive oil-agar plates suggesting hydrolysis of long-chain triglycerides. Three more mutants M2a, M4a and M4b also showed hydrolysis of olive oil on emulsified agar plates. Interestingly these mutants all contain the R37A mutation (Figure S7). Every mutant that contains the R37A mutation shows the emergence of this new activity suggesting that mutation at position 37 is essential for this new catalytic ability. A similar conclusion was drawn for the rumen extracted esterase R.34 in which the formation of a single salt bridge turn the enzyme into a true lipase without modification of the shape size or hydrophobicity of the substrate-binding pocket that are usually considered to be essential for chain-length specificity [23]. The exact mechanism by which removing of the arginine residue at position 37 lead to the emergence of this new function in EstGtA2 is under investigation.

Modification of the (i - 2, i - 4) interloop salt bridge

A highly conserved interloop salt bridge was identified in distantly related bacterial lipolytic enzymes as discussed above. The bridge was modulated by replacing His222 by Arg which turns the bridge to Asp194-Arg222, a pairing similar to the one found in MGL H-257 and in MGL from *G. kaustophilus* from the N' subfamily (Figure 5), so this mutation should be easily accommodated. The effect of this substitution on thermal stability was investigated further using CD spectroscopy. The pH-dependent thermal unfolding profile of the wild type and of the H222R mutant was recorded from pH 4 to 10. Results showed that the H222R mutation contributed to increase the apparent T_m value under alkaline pH (at pH 10) by +2.1°C (Figure 11). Activity measurements confirmed that H222R was active over the same pH range as the wild type EstGtA2 (pH 4 to 10) with optimal activity at pH 8 (data not shown). However, at pH 10, the H222R mutant retained more residual activity at room temperature after heating, with 89% left compared to 46% for the wild type enzyme (after heating at 57°C) and 74% left compared to 22% for the wild type after heating at

60°C (Figure 12). These results can be explained by the difference in pK_a between the histidine and the arginine, something which will determine the pH under which the electrostatic interaction would be optimally formed. As the pH increases to alkaline conditions, the Asp194-His222 salt bridge weakens as His222 loses its extra proton. Under these conditions, the mutant H222R was stabilised by an additional salt bridge compared to wild type enzyme. Despite that the H222R substitution increased T_m under alkaline conditions (pH 10), it led to a decrease in T_m at pH 5 by - 1.7°C.

Discussion

In our group we have elected to focus on closely related proteins for which sufficient information (evidence of expression, thermal stability and other enzymatic properties) is available. Because of this relatedness, we believe that sequence-properties relationships inside the boundaries of a given protein family can be used successfully for predicting or controlling properties on the basis of DNA sequences analysis. Lipolytic enzymes from the recently identified N' subfamily shares a very high degree of residue identity, yet display differences in stability and other properties. Consequently, they appear as ideal subjects for investigating the role of salt bridge residues in determining stability in this particular enzyme family.

A 15th family of bacterial lipolytic enzymes

We proposed the 15th family (family XV) which correspond to abH11.03 on the basis of a distinct phylogenetic cluster (< 32% with family XIII), (Figure S8), the architecture of the cap domain (α/β cap) compared with the three α -helices cap of Est30-like proteins and suggest its separation and identification of a subfamily named N' on the basis of a distinctive salt bridges pattern and on the polarity of a conserved interloop salt bridge in the active site. Similarly a distinctive pattern is conserved among the Est30-like group forming a subfamily within family XIII (N subfamily).

The proposed N' subfamily contain lipolytic enzymes, which are typically ~ 250 residues carboxylesterases or monoacylglycerol lipases [7]. Structure comparison showed that the residues of the binding pocket are well conserved among EstGtA2 and MGL H-257, suggesting that other residues have an impact on specificity or other enzymatic properties. N' enzymes have an additional domain over the active site (similar to the lid of lipases); they process long chain monoacylglycerol, but do not hydrolyze or synthesize long-chain triglycerides. Hydrolysis of long-chain triglycerides is a particular property of true lipases (EC 3.1.1.3) for which more elaborated lid structures allow for opening at the lipid-water interface, a phenomenon called interfacial activation. The architecture of the cap domain for N' enzymes involves residues 125 to 163 (Figures 2-4). The recently identified LipS enzyme which shares many features with N' enzymes, also has a similar cap (named here α/β cap). Other α/β hydrolases from distantly related organisms exhibit similar architecture in the cap domain. Interestingly, a cap architecture similar to EstGtA2, MGL H-257 or LipS was found in the crystal structure of distant α/β hydrolases including Est1E from *Butyrivibrio proteoclasticus* (2WTM), *Vibrio harveyi* thioesterase (1THT), esterase Rv0045c from *Mycobacterium tuberculosis* (2P2M), *Actinida eriantha* carboxylesterase AeCXE1 (2O7R), and *Streptococcus pneumoniae* EstA (2UZ0).

The corresponding cap is not found in closely related Est30 from *G. stearothermophilus* (PDB no. 1TQH). In this case the cap is formed by three α -helices (D1' D2' B1'), involving residues 123 to 157 (Figures 3 and 4). The cap (this cap version will be referred to as α cap) architecture of the *G. stearothermophilus* Est30 (1TQH), which involves three helices, is also found in several distantly related carboxylesterases including: *Staphylococcus aureus* MenH (2XMZ), *Escherichia coli* ybfF (3BF7) and BioH (1M33), *Streptomyces aureofaciens* bromoperoxidase (1BRO), *Amycolatopsis mediterranei* thioesterase (3FLA) and *Agrobacterium radiobacter* epoxyde hydrolase (1EHY). Cap structure comparison supports the divisions proposed in phylogenetic analysis (Figure 5).

We found here that on the basis of the conserved salt bridges content, the cap structure, and on the polarity of the interloop salt bridge, N'enzymes clearly distinguishes from the other members of abH11.03 (or proposed family XV). LipS properties are shared by enzymes from both N' and N subfamilies: it has the α/β cap typical for N', but it has the interloop polarity of N (family XIII) and lacks the five bridges exclusive to N'. On the basis of our analysis of available sequences it appears that the N cluster includes Est30-like enzymes from family abH11.01 [9]. Similarly to N', the N cluster forms a subgroup based on a distinctive salt bridges pattern.

Based on an earlier classification of α/β hydrolase enzymes (Lipase Engineering Database) as suggested by Fischer and Pleiss [9], carboxylesterases are grouped in the abH11 family. In this classification, the carboxylesterase Est30 (1TQH) belongs to abH11.01 while EstGtA2 belongs to the abH11.03 subfamily (Figure 5). In addition, the recently suggested LipS family [8], appears to belong to abH11.03. Therefore, the abH11.01 family (Fischer) and XIII [4] appear to be equivalent. The N' cluster might be considered as a subfamily of abH11.03 (family XV). We propose to extend the Arpigny and Jaeger's classification by introducing the new family XV (corresponding to abH11.03). The assignment of various clusters using the various classification schemes mentioned here is shown in Figure 6. The particular and exclusive salt bridges content and specific polarity of the interloop salt bridge are characteristics of the N' subfamily.

Role of key salt bridges in thermostability of EstGtA2

The demonstration that thermal denaturation (and its reversibility) was pH-dependent for EstGtA2 strongly supports a role for ionisable side chains in its stabilisation [19, 20]. Thus, we chose to explore the role of arginine and lysine side-chains involved in the formation of the five conserved salt bridges in the N' subfamily. Alanine scanning was chosen among other methods because we wanted to use a combinatorial approach, which cannot be reconciled with sophisticated approaches such as double-mutant cycle analysis (DMC) [45-46]. The substitution of alanine for arginine or lysine can be compared to "shaving" the side-chains. In this respect, one has to keep

in mind that “alanine-shaving” of a particular side-chain removes electrostatic interactions as well as hydrophobic packing [47, 48]. Every alanine substitution at Arg of Lys residues involved in salt bridges was found to decrease the stability of EstGtA2 at high temperature. The decrease in T_m was generally more pronounced for multiple mutations than for single ones. In all cases except one, folding and activity were observed at room temperature.

We show that when a particular residue is mutated (R37A), low temperature activity is increased, which might be due to change in enzyme flexibility, and concomitant favouring of higher activity at lower temperature [13]. This mutation also leads to shift the optimal temperature from 50 to 55°C compared to wild type. Moreover, this mutation was found to confer activity on long-chain triglycerides. Hydrolysis of long-chain triglycerides was detected for every mutant containing the R37A mutation. The mechanism by which this mutation changes specificity is under investigation.

Removal of bulky Arg led to limited changes in EstGtA2 tertiary structure packing, as indicated by near UV CD. Results from near and far-UV CD and fluorescence measurements indicate that EstGtA2 accommodated Ala substitutions without major structure modification in all but the M4c mutant. Thus, if we exclude the mutant M4c from the analysis, the five bridges or residues mutated do not appear to play a key role in specifying folding. It appears that their accumulation, and not their specific location or interactions, lead to loss of EstGtA2 stability and prevent correct folding of the enzyme in the case of M4c. The estimated conformational stability ($\Delta G_{25^\circ\text{C}}$) for the wild type EstGtA2 (while partly reversible) at pH 8 is 11.8 kcal/mol, while for M4a is 4.9 kcal/mol. Despite a considerable reduced stability ($\Delta\Delta G_{25^\circ\text{C}} = - 6.9$ kcal/mol), M4a remains active at low temperature. In the case of M4c, it appears that the ability of EstGtA2 to accept mutations of large and positively charged residues has been exceeded. What exact mutation(s) led to loss of compactness and activity at room temperature for this mutant is not clear. A similar conclusion was reached for indole-glycerol-phosphate synthase [12].

The extent to which the P22 Arc repressor can tolerate alanine substitutions has been reported. Interestingly, twenty five neutral positions were found and among these, the combination of 15 multiple-alanine substitutions can be tolerated without changing its conformation and/or general properties. However, the majority of the positions that do not tolerate alanine substitution involve residues forming the hydrophobic core of the protein or forming buried hydrogen bonds and salt bridges [49-52]. It was also reported that two buried salt bridges have a considerable impact in the folding pathway and the stability of Barnase [53, 54]. This suggests that a limited number of the residues of a protein may determine its folding and its stability. Although that multiple salt bridges are directly related to thermostability of proteins from thermophiles, it is not clear how and to which extent they can play a determinant role in the folding process [55, 56]. Distinct patterns of salt bridges were observed in closely-related proteins of the same fold. The role that may play these different salt bridge patterns in influencing protein folding and/or stabilizing the native conformation need to be explored.

The second part of this study focused on the interloop salt bridge D194-H222 (in EstGtA2). This particular bridge is found in N, N' families and in some lipolytic enzymes that have not been assigned to any of the families of the Arpigny and Jaeger classification (I to XIV). In particular we focused on His222 and the possible importance of pK_a value in determining pH dependence of enzymatic properties. As suggested by multiple alignments and phylogeny presented here, this bridge is very tolerant to mutation and accepted all five "usual" pH-responsive salt bridge-forming residues (Arg, Lys, His, Glu and Asp). Furthermore, the bridge can have any polarity in distantly related enzymes: in the N' subfamily it is in opposite orientation compared to the corresponding bridge for enzymes from the N subfamily and close homolog LipS. The residue identity is switched in some cases. Considering this apparent flexibility and possible importance for enzymatic properties, we replaced the His residue found at position 222 in EstGtA2 for Arg. We found that His provided more stability at pH 5 than Arg at the same position. The fact that Arg did not confer the same stabilisation that His at the same position did below pH 5 might be explained by the different contributions of side-chain non-ionic interactions with the protein, as discussed previously [17].

Nevertheless, at pH 10, Arg provided a three-fold improvement in thermostability at 60°C. The high pK_a of Arg residue made that it should remain positive at higher pH, and thus maintain the coulombic interaction with D194. Is the proximity of the active site the explanation for this impact remains to be demonstrated, but clearly, the interloop salt bridge, common to families N (XIII or abH11.01) and N' (XV or a sub-classification of abH11.03) is a determinant of thermostability.

Conclusion

We identified five salt bridges that are a hallmark of lipolytic enzymes from the N' subfamily (family XV), suggested by Charbonneau *et al.*, 2010. We showed that the five positively charged residues involved in these bridges provide important stability and are essential for activity at high temperature. These key salt bridge-forming residues can be used for fine-tuning activity-temperature relationship. Despite their strict conservation in enzymes from the N' subfamily, up to four salt bridge disruptions can be tolerated before EstGtA2 becomes unable to adopt its native fold. We also identified co-evolved residues within family XV and XIII that form a particular interloop salt bridge near the active site. Mutation at this interloop position in EstGtA2 (D194/H222), changing the His residue at position 222 for Arg, improved enzyme thermostability several fold under alkaline pH. Our study suggests primary targets for the optimization of EstGtA2 and enzymes from the N' subfamily for specific biotechnological applications. These conserved salt bridge-forming residues may be useful sequence indicators for the assignation of newly discovered N'-related enzymes in the future.

Supporting information

Figure S1. Structural model of EstGtA2. Structural alignment of EstGtA2 model and X-ray crystal structure of MGL H-257 (left) and with a 90° rotation (right). Conserved salt bridges studied are shown. Conserved residues of the binding site are shown in orange.

Figure S2. Amino acids substitution between EstGtA2 and MGL H-257. The 27 out of 249 residues that are different in EstGtA2 compared to MGL H-257 (89% identity) are shown in red (EstGtA2) and orange (MGL H-257).

Figure S3. Refinement statistics for the EstGtA2 model. Ramachandran plot for the EstGtA2 model based on the X-ray crystal structure of MGL H-257 (PDB no. 3RM3).

Figure S4. The E3-R54 salt bridge. The first 86 residues from the N-terminal end are shown. The E3-R54 salt bridge links the N-terminal end of EstGtA2 to the core. A hydrogen bond between the R54 guanidinium and the oxygen of H58 and A11 backbone is predicted. In addition the R54 side-chain would form a hydrophobic cluster with two prolines (P13 before strand β 2 and P33 after helix α 1).

Figure S5. Salt bridges interactions. The predicted interactions for the following salt bridges: E3-R54 and E12-R37 (A), the E66-R140 (B), D124-K178 (C), D205-R220 and D194-H222 (D).

Figure S6. Unfolding free energy for EstGtA2 and mutants. The unfolding free energy (ΔG) as function of temperature (K) is shown. The melting temperatures (T_m) are determined at $\Delta G = 0$.

Figure S7. Hydrolysis of long-chain triglycerides by R37A. The wild type EstGtA2 and mutants were deposited (10 μ g) onto emulsified olive oil-agar plate containing rhodamine 0.001%. The hydrolysis of TAG released free fatty acids and the activity was detected under UV-illumination at 302 nm.

Figure S8. The new family XV. Phylogenetic tree showing the relationship between identified bacterial lipolytic enzyme families (I-XV). The new 15th family is shown in bold.

File S1. Supplementary tables. Table S1 Primers used for cloning and directed mutagenesis. Table S2 Salt bridges studied. Table S3 Distinctive salt bridges composition between the N', LipS and N clusters.

Acknowledgements

The authors are grateful to M.H. Hefford and P. Pawelek for their editorial contribution to different sections of this report.

References

1. Ollis DL, Cheah E, Cygler M, Dijkstra B, Frolow F *et al.* (1992) The alpha/beta hydrolase fold. *Protein Eng* 5(3): 197-211.
2. Nardini M, Dijkstra BW (1999) α/β hydrolase fold enzymes: the family keeps growing. *Current Opinion in Structural Biology* 9: 732-737.
3. Arpigny JL, Jaeger K-E (1999) Bacterial lipolytic enzymes: classification and properties. *Biochem J* 343: 177-183.
4. Rao L, Xue Y, Zheng Y, Lu JR, Ma Y (2013) A novel alkaliphilic *Bacillus* esterase belongs to the 13th bacterial lipolytic enzyme family. *PLoS ONE* 8(4): e60645. doi:10.1371/journal.pone.0060645.
5. Montoro-Garcia S, Martinez-Martinez I, Navaro-Fernandez J, Takami H, Garcia-Carmona F *et al.* (2009) Characterization of a novel carboxylesterase from *Geobacillus kaustophilus* HTA426 shows the existence of a new carboxylesterase family. *J Bacteriol* 191(9):3076-3085.
6. Liu P, Wang YF, Ewis HE, Abdelal AT, Lu CD *et al.* (2004) Covalent reaction intermediate in crystal structure of the *Geobacillus stearothermophilus* carboxylesterase Est30. *J Mol Biol* 342: 551-561.
7. Charbonneau DM, Meddeb-Moulehi F, Beauregard M (2010) A novel thermostable carboxylesterase from *Geobacillus thermodenitrificans*: Evidence for a new carboxylesterase family. *J Biochem* 148: 299-308.
8. Chow J, Kovacic F, Dall Antonia Y, Krauss U, Fersini F, *et al.* (2012) The Metagenome-Derived Enzymes LipS and LipT Increase the Diversity of Known Lipases. *PLoS ONE* 7(10): e47665. doi:10.1371/journal.pone.0047665.
9. Fischer M, Pleiss J (2003) The lipase engineering database: a navigation and analysis tool for protein families. *Nucleic Acids Res* 31: 319-321.
10. Bornscheuer UT (2002) Microbial carboxyl esterases: classification, properties and application in biocatalysis. *FEMS Microbiol Rev* 26: 73-81.
11. Vetriani C, Maeder DL, Tolliday N, Yip KS, Stillman TJ, *et al.* (1998) Protein thermostability above 100 degrees C: a key role for ionic interactions. *Proc Natl Acad Sci USA* 95: 12300-12305.

12. Knöchel T, Pappenberger A, Jansonius JN, Kirschner K (2002) The crystal structure of indoleglycerol γ -phosphate synthase from *Thermotoga maritima*. Kinetic stabilization by salt bridges. *J Biol Chem* 277: 8626-834.
13. Bae E, Phillips G. N (2004) Structures and analysis of highly homologous psychrophilic, mesophilic, and thermophilic adenylate kinases. *J Biol Chem* 279: 28202–28208. doi: 10.1074/jbc.M401865200.
14. Kumar S, Tsai C-J, Nussinov R (2000) Factors enhancing protein thermostability. *Prot Eng* 13: 179-191.
15. Strop P, Mayo SL (2000) Contribution of surface salt bridges to protein stability. *Biochemistry* 39(6): 1251-1255.
16. Takano K, Tsuchimori K, Yamagata Y, Yutani K (2000) Contribution of salt bridges near the surface of a protein to the conformational stability. *Biochemistry* 39(40): 12375-12381.
17. Makhatadze GI, Loladze VV, Ermolenko DN, Chen X, Thomas ST (2003) Contribution of surface salt bridges to protein stability: guidelines for protein engineering. *J Mol Biol* 237(5): 1135-1148.
18. Kumar S, Nussinov R (1999) Salt bridges stability in monomeric proteins. *J Mol Biol* 293(5): 1241-1255.
19. Mulkerrin MG, Wetzel R (1989) pH-dependence of the reversible and irreversible thermal denaturation of γ interferon. *Biochem* 28: 6556-6561.
20. Perutz MF (1978) Electrostatic effects in proteins. *Science* 201: 1187-1191.
21. Li WF, Zhou XX, Lu P (2005) Structural features of thermozymes. *Biotechnology Advances*. 23: 271-281.
22. Vielle C, Zeikus GJ (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanism for thermostability. *Microbiol Mol Biol Rev* 65(1): 1-43.
23. Reyes-Duarte D, Polaina J, Lopez-Cortez N, Alcalde M, Plou FJ, Elborough K *et al.* (2005) Conversion of a carboxylesterase into a triacylglycerol lipase by a random mutation. *Angew Chem Int Ed Engl* 44(46): 7553-7557.
24. Kumar S, Nussinov R (2002) Relationship between ion pair geometries and electrostatic strengths in proteins. *Biophys J* 83(3): 1595-1612.

25. Dadarlat VM, Post CM (2008) Contribution of charged groups to the enthalpic stabilization of the folded states of globular proteins. *J Phys Chem B* 112(19): 6159-6167.
26. Jelsarov I, Karchikoff A (2009) Defining the role of salt bridges in protein stability. *Methods. Mol. Biol* 490: 227-260.
27. Yennamalli RM, Rader AJ, Wolt JD, Sen TZ (2011) Thermostability in endoglucanase is fold-specific. *BCM Structural Biology* 11(10). doi: 10.1186/1472-6807-11-10.
28. Johnson WC (1999) Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins* 35(3): 307-312.
29. Sreerama N, Woody RW (2000) Estimation of protein secondary structures from circular dichroism spectra: Comparison of CONTIN, SELCON and CDSSTR with an expended reference set. *Analytical Biochemistry* 287: 252-260.
30. Pace CN, Shirley BA, Thompson JA (1989) Measuring the conformational stability of a protein. In: Creighton TE, editors. *Protein structure-A practical approach*. Oxford University Press. Pp. 311-330.
31. Boratyn GW and others (2013) BLAST: a more efficient report with usability improvements. *Nucleic Acids Res* doi: 10.1093/nar/gkt282
32. Benson AD, Cavanaugh M, Clark K, Mizrachi-Karsch I, Lipman DJ *et al.* (2013) GenBank. 41: doi:10.1093/nar/gks1195
33. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22(22): 4673-4680.
34. Gouet P, Courcelle E, Stuart DI, Métoz F (1999) ESript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* 15(4): 305-308.
35. Tamura K, Peterson D, Peterson N, Stecher G, Nei M *et al.* (2011) MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol* 28(10): 2731-2739.
36. Saitou N, Nei M (1987) The neighbour-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol* 4(4): 406-425.

37. Lambert C, Léonard N, De Bolle X, Depiereux E (2002) ESyPred3D : Prediction of protein structures. *Bioinformatics* 18: 1250-1256.
38. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK - a program to check the stereochemical quality of protein structures. *J. App. Cryst.* 26: 283-291.
39. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* 22: 195-201.
40. Constantini S, Colonna G, Facchiano AM (2008) ESBRI: A web served for evaluating salt bridges in proteins. *Bioinformation* 3(2): 137-138.
41. Li H, Robertson AD, Jensen JH (2005) Very fast empirical prediction and rationalization of protein pKa values. *PROTEINS* 61: 704-721.
42. Kitaura S, Suzuki K, Imamura S (2001) Monoacylglycerol lipase from moderately thermophilic *Bacillus* sp. Strain H-257: Molecular cloning, sequencing and expression in *Escherichia coli* of the gene. *J Biochem* 397-402.
43. Imamura S, Kitaura S. (2000) Purification and characterization of a monoacylglycerol lipase from the moderately thermophilic *Bacillus* sp. H-257. *J Biochem* 419-425.
44. Rengachari S, Bezerra GA, Roegler-Berket L, Gruber CC, Sturm C *et al.* (2012) The structure of monoacylglycerol lipase from *Bacillus* sp H-257 reveals unexpected conservation of the cap architecture between bacterial and human enzymes. *Biochimica et Biophysica acta* 1012-1021.
45. Carter PJ, Winter G, Wilkinson AJ, Fersht AR (1984) The use of double mutants to detect structural changes in the active site of tyrosyl-tRNA synthetase (*Bacillus starothermophilus*). *Cell* 3: 835-840.
46. Morrison KL, Weiss GA (2001) Combinatorial alanine-scanning. *Curr Opin Struct Biol* 5: 302-307.
47. Yang Y, Yu Y, Cheng J, Liu Y, Liu D-S *et al.* (2012) Highly conserved salt bridge stabilizes rigid signal patch at extracellular loop critical for surface expression of acid-sensing ion channels. *J Bio Chem* 287(18): 14443-14445.
48. Yang J, Wu J, Steichen JM, Kornev AP, Deal MS *et al.* (2012) A conserved Glu-Arg salt bridge connects coevolved motifs that define the eukaryotic kinase fold. *J Mol Biol* 415(4): 666-679.

49. Milla ME, Brown, BM, Sauer RT (1994) Protein stability effects of a complete set of alanine substitutions in Arc repressor. *Nat. Struct. Biol* 1(8): 518-523.
50. Brown BM, Sauer RT (1998) Tolerance of Arc repressor to multiple-alanine substitutions. *Proc. Natl. Acad. Sci. USA* 96: 1983-1988.
51. Waldburger CD and Schildbach JF, Sauer RT (1995) Are buried salt bridges important for protein stability and conformational specificity. *Nat. Struct. Biol* 2: 122-128.
52. Sauer RT, Waldburger CD, Brown BM, Schildbach JF (1996) Sequence determinant of folding and stability for the P22 Arc repressor. *FASEB* 10: 42-48.
53. Tissot AC, Vuilleumier S, Fersht AR (1996) Importance of two buried salt bridges in the stability and folding pathways of Barnase. *Biochemistry* 35: 6786-6794.
54. Oliveberg M, Fersht AR (1996) A new approach to study of transient protein conformations: The formation of a semiburied salt link in the folding pathway of Barnase. *Biochemistry* 35: 6795-6805.
55. Donald JE, Kulp DW, DeGrado DW (2011) Salt bridges: Geometrically specific, designable interactions. *Proteins* 79(3): 898-915.
56. Karshikoff A, Ladenstein R (2001) Ion pairs and the thermotolerance of proteins from hyperthermophiles: a “traffic rule” for hot roads. *Trends Biochem. Sci* 26: 550-556.

Table 1. Impact of salt bridges disruption on EstGtA2 activity and stability. Melting temperatures and specific activities were measured at pH 8. The amount of *p*Np-octanoate hydrolyzed as a function of time. Specific activities ($\mu\text{mol min}^{-1} \text{mg}^{-1}$) are reported in percentage (relative to the wild type in respective temperature). The initial rates (V_0) were measured under saturation and than reflect k_{cat}/K_M . Each thermal denaturation curves or rate of hydrolysis were collected at least three times. The T_m and activity values are average of three different experiments. Standard deviations based on respective triplicates are shown.

Protein	Mutations	T_m (°C)	ΔT_m (°C)	Activity (%) ^a	Activity (%) ^b
WT	-	64.2 ± 0.11	-	100	100
M1a	R37A	58.6 ± 0.31	$- 5.6 \pm 0.21$	123 ± 3.6	143 ± 3.7
M1b	R54A	55.6 ± 0.25	$- 8.6 \pm 0.18$	68.1 ± 2.3	27.2 ± 5.1
M1c	R140A	56.6 ± 0.23	$- 7.6 \pm 0.17$	108 ± 5.1	112 ± 2.1
M1d	K178A	55.3 ± 0.23	$- 8.9 \pm 0.17$	108 ± 2.4	103 ± 2.3
M1e	R220A	55.4 ± 0.15	$- 8.8 \pm 0.08$	108 ± 0.8	68 ± 0.9
M2a	R37A/R220A	55.6 ± 0.18	$- 8.6 \pm 0.15$	n.d.	n.d.
M2b	R140A/R220A	54.1 ± 0.19	$- 10.1 \pm 0.15$	60.2 ± 4.5	92.1 ± 3.6
M2c	K178A/R220A	50.6 ± 0.29	$- 13.6 \pm 0.20$	112 ± 3.3	76.2 ± 4.1
M3a	R54A/R140A/R220A	53.9 ± 0.31	$- 10.3 \pm 0.22$	71.3 ± 2.7	4.56 ± 2.1
M3b	R54A/K178A/R220A	52.3 ± 0.17	$- 11.9 \pm 0.14$	108 ± 2.7	8.31 ± 1.7
M3c	R140A/K178A/R220A	50.4 ± 0.10	$- 13.8 \pm 0.09$	47.3 ± 3.3	0
M4a	R37A/R54A/R140A/R220A	53.2 ± 0.23	$- 11.0 \pm 0.17$	47.2 ± 2.8	0
M4b	R37A/R140A/K178A/R220A	49.9 ± 0.30	$- 14.3 \pm 0.21$	79.1 ± 1.9	14.5 ± 2.3
M4c ^U	R54A/R140A/K178A/R220A	-	-	0	0
WT ^c	-	54.5 ± 0.21	-	100	100
ISB ^c	H222R	56.6 ± 0.18	$+ 2.5 \pm 0.24$	98.2 ± 0.8	254 ± 5.1

U. Unfolded and completely inactivated at 25°C.

a. Activity relative to wild type at 25°C.

b. Activity relative to wild type at 60°C.

c. Proteins properties recorded at pH 10 using 20 mM CAPS buffer.

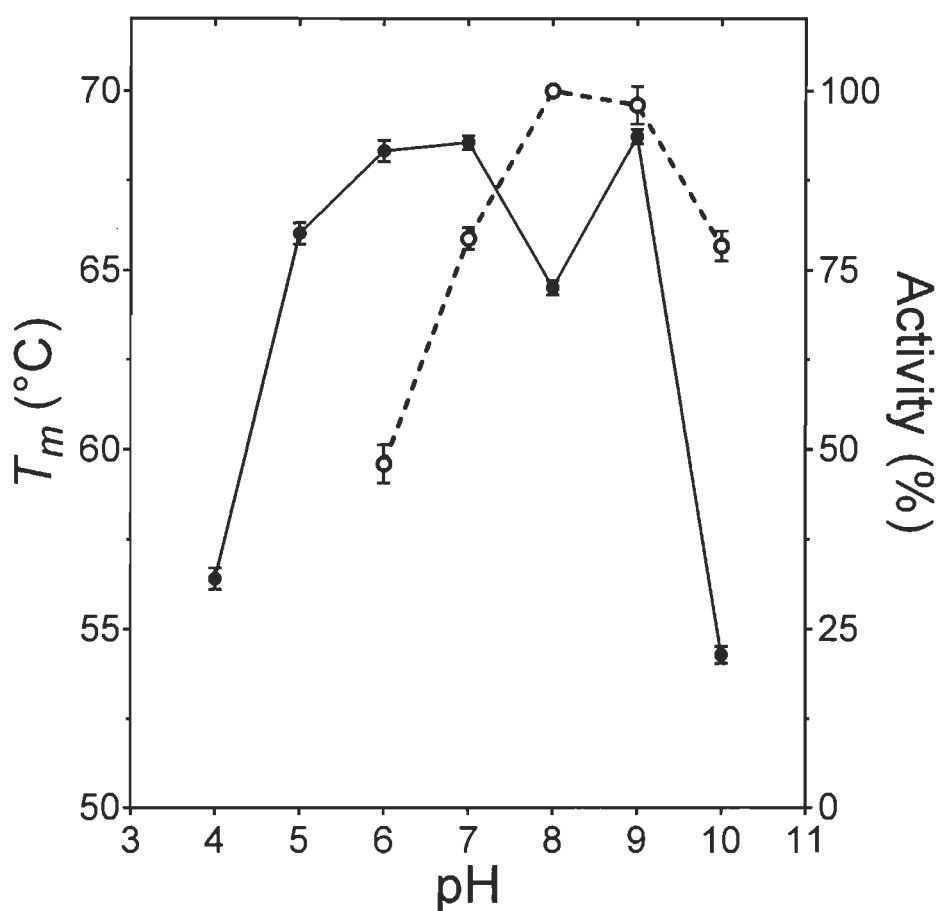


Figure 1. Activity and stability dependence on pH for EstGtA2. The apparent melting temperatures (derived from CD) as function of pH are shown as closed circles. Open circles show relative activity (hydrolysis of *p*NP-octanoate) at different pH values at 50°C (optimal temperature). Standard deviations for T_m ranged between 0.1-0.3°C and do not exceeded 5% for activity. Buffers used were sodium citrate/citric acid (pH 4-5), sodium phosphate (pH 6-8), Tris-HCl (pH 9) and CAPS/NaOH (pH 10-11).

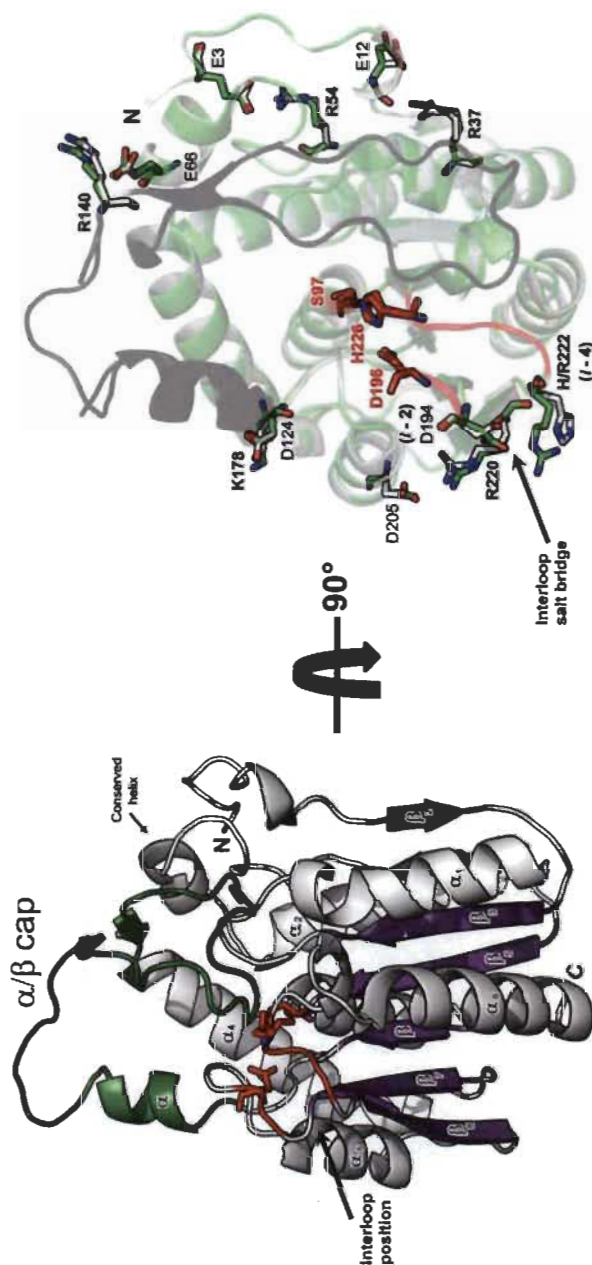


Figure 2. Structure model of EstGtA2. Ribbon structure of EstGtA2 based on the X-ray crystal structure of MGL H-257 (left). The beta sheet is shown in purple and the seven strands are identified as $\beta 2$ - $\beta 8$, alpha helices are in grey and identified as $\alpha 1$ - $\alpha 6$. The cap domain (residues 125-161) is shown in green. The catalytic triad S97, D196, H226 is shown as red sticks. Right panel shows a structural alignment of EstGtA2 model (white) and MGL H-257 crystal structure (3RM3) in light green. The five salt bridges studied are shown (E3-R54, E12-R37, E66-R140, D124-K178 and D205-R220) as sticks. The interloop salt bridge conserved in ($i - 2$, $i - 4$) from the catalytic Asp and His residues respectively is identified by an arrow (D194-H222 in EstGtA2).

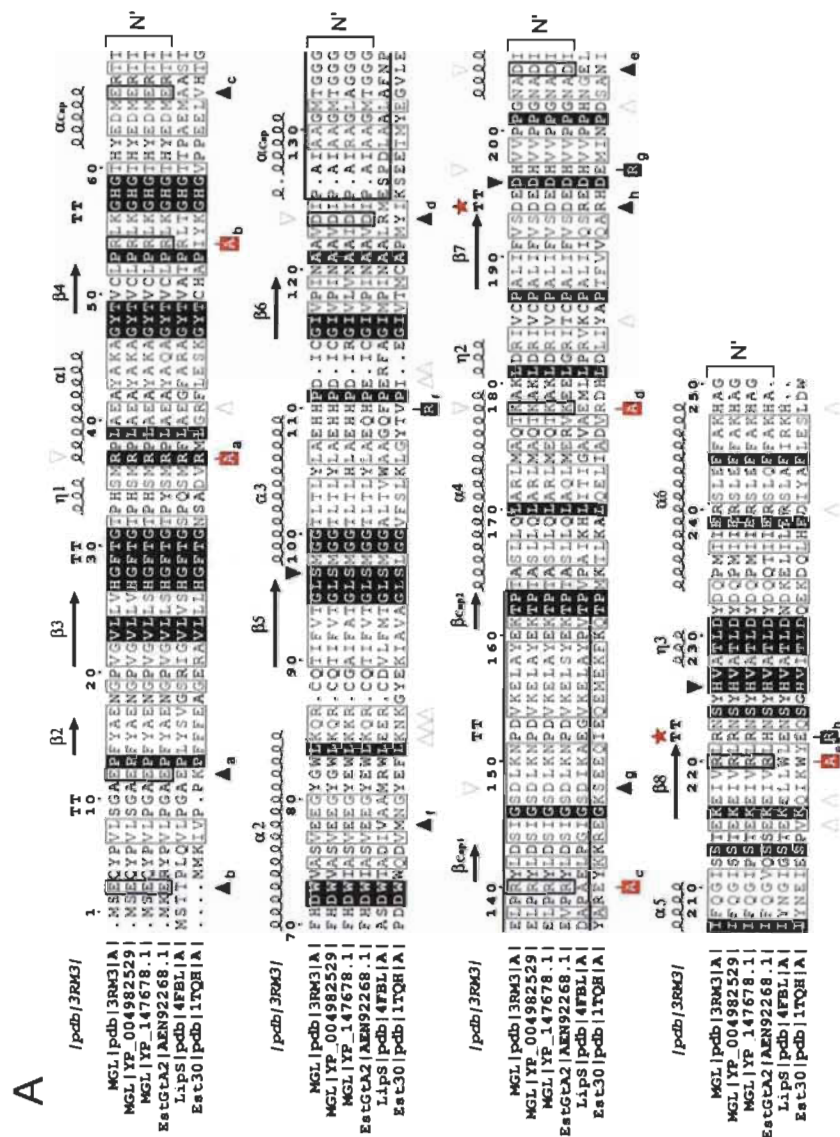


Figure 3A. Distinctive salt bridges patterns within family XV and XIII. Panel A: Multiple sequence alignment for N' subfamily enzymes (MGL from *Bacillus* sp. H-257, *G. thermoleovorans*, *G. kaustophilus* and EstGtA2 from *G. thermodenitrificans*) and representative of the LipS (4FBL) and N subfamily (1TQH). The secondary structure elements from MGL H-257 (3RM3) are shown on top. The cap domain is boxed (thick line). Residue numbering is based on EstGtA2 and MGL H-257. The residues of the catalytic triad are identified by a triangle above the MSA. The seven N' conserved salt bridges: E3-R54, E12-R37, E66-R140, H110-E78, D124-K178, H197-D148 and D205-R220 are identified by close triangles below the MSA with same letter. Residues shaved by alanine-scanning are identified by a red box below the MSA. Open triangles refer to LipS salt bridges pattern. The tolerant interloop salt bridge located in (*i*-2, *i*-4) from the catalytic Asp and His respectively is identified with a red star above the MSA. Mutations by Ala or Arg are indicated by a boxed A or R.

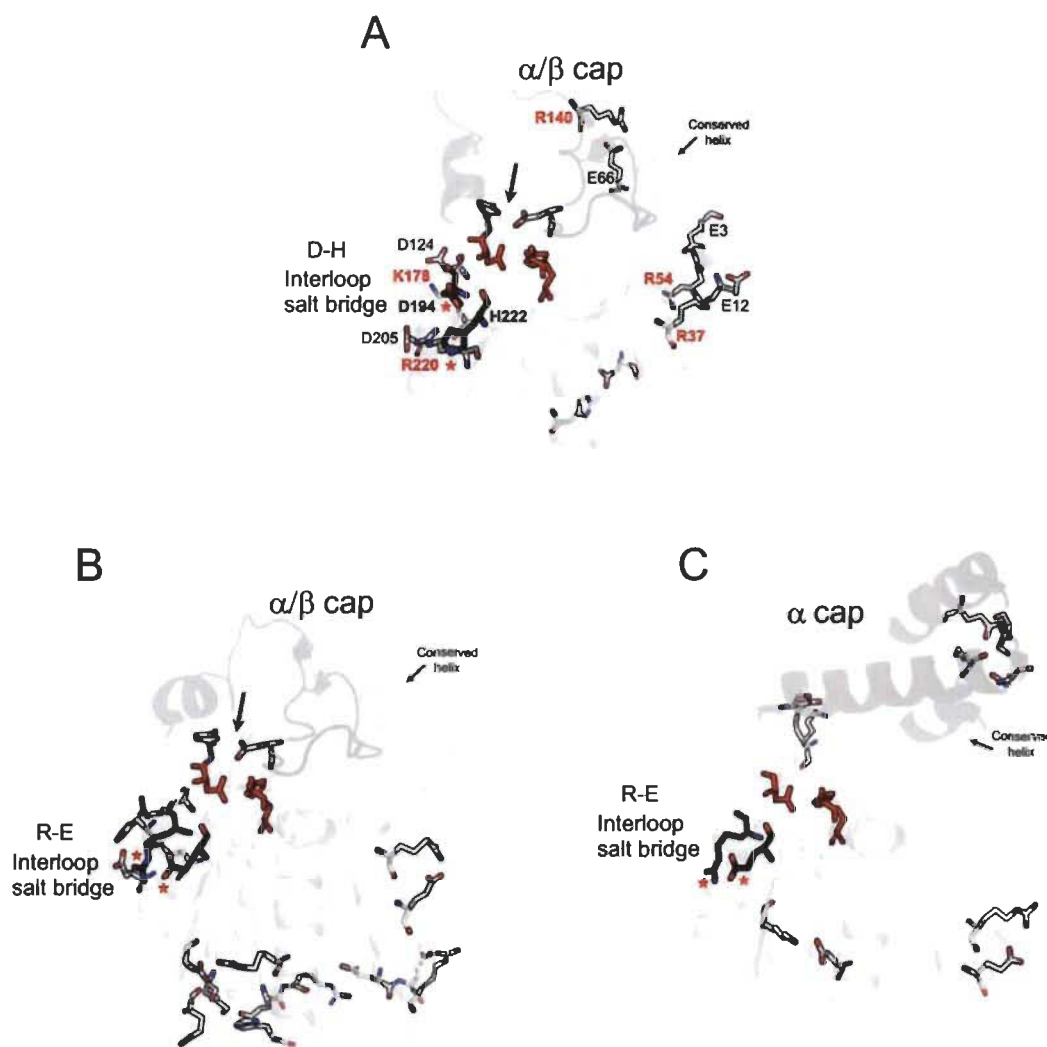
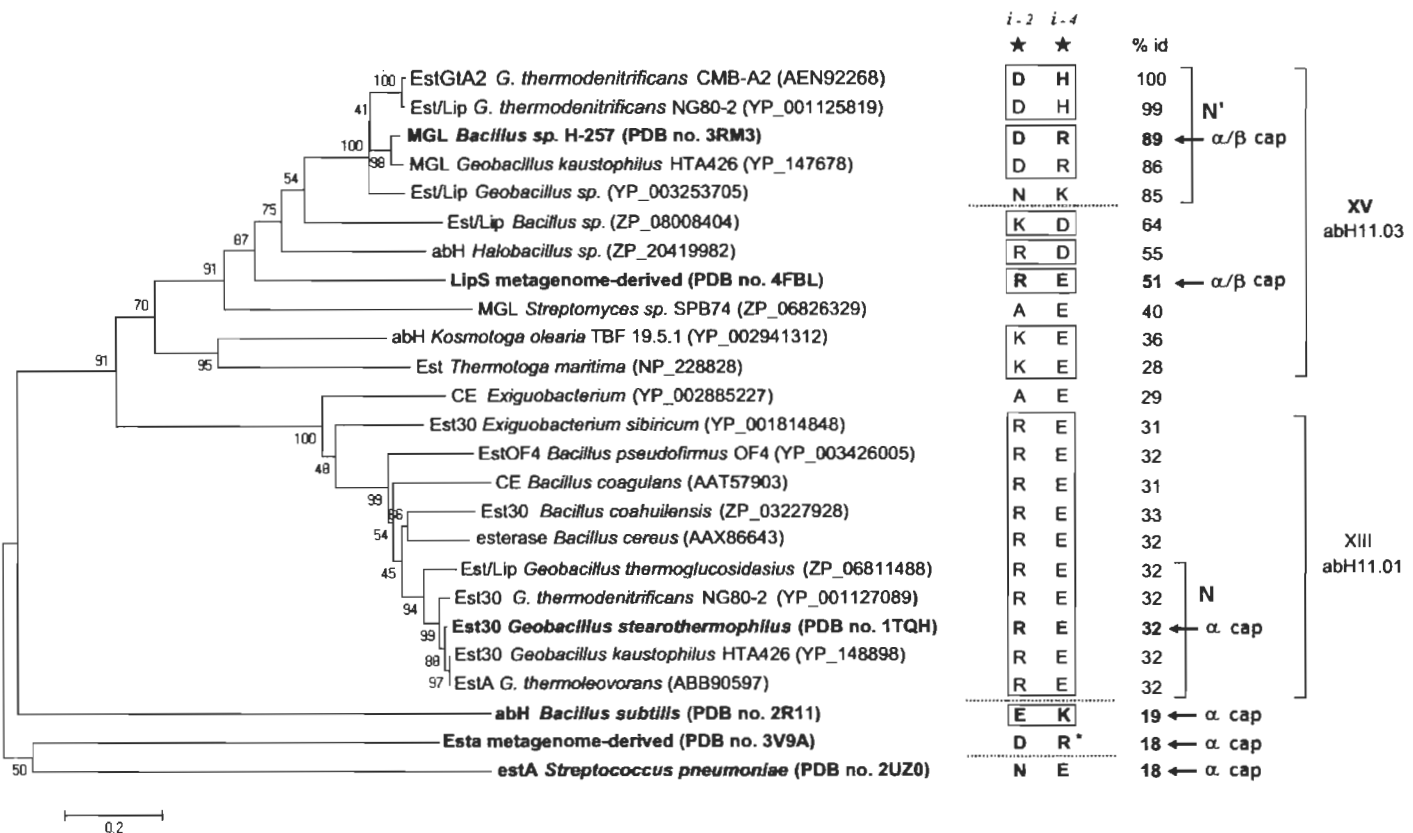


Figure 4. Distinctive salt bridge patterns conserved in related lipolytic enzymes. Structural comparison of the salt bridges content for EstGtA2 (A) LipS (4FBL) (B) and Est30 (1TQH) (C) is shown. At variance with the conserved interloop bridge, the 5 selected salt bridges in EstGtA2 are absent in LipS- and N-related enzymes. Residues exclusive to the N' subfamily and studied by alanine-scanning mutagenesis are shown in red. Structure of the cap domain is shown in dark gray. The conserved helix of the cap between the three structures is identified. LipS structure share features from N' and from N subfamilies. The arrow shows salt bridge conserved between EstGtA2 and LipS. The residues forming the interloop salt bridge are shown and identified on the respective structures.



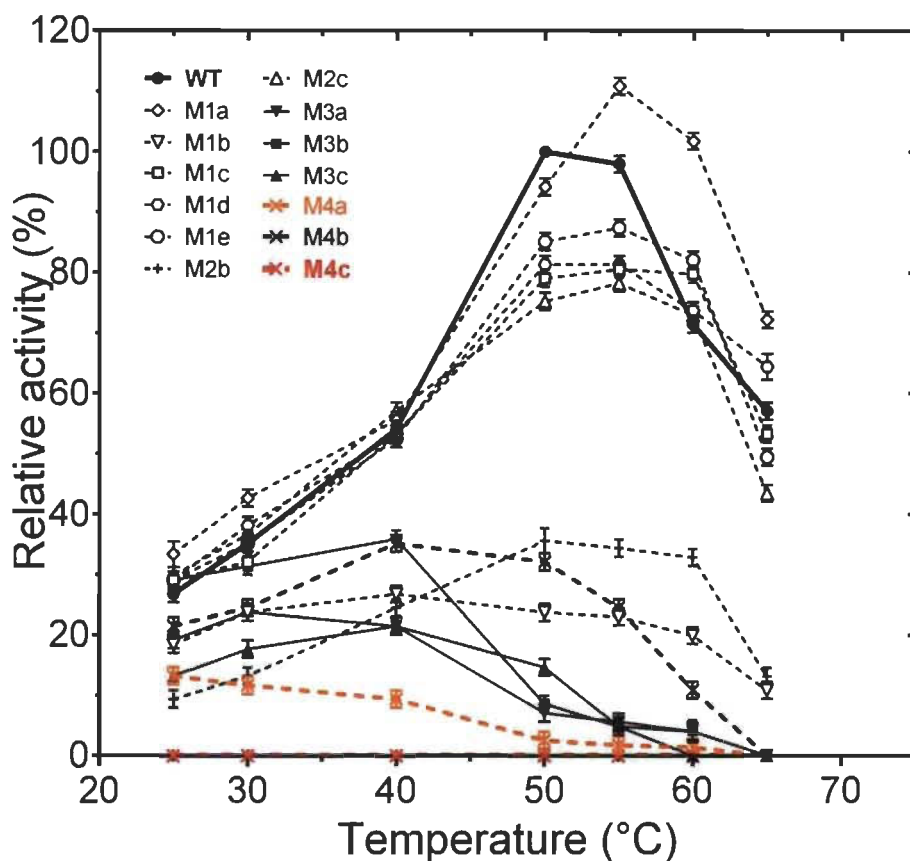


Figure 6. Activity profile for wild type and mutants. Specific activity ($\mu\text{mol min}^{-1} \text{mg}^{-1}$) was measured from 25 to 65°C for the wild type and mutants in 20 mM sodium phosphate pH 8 using *p*NP-octanoate as substrate. The initial rates (V_0) were measured below substrate saturation and than reflect k_{cat}/K_M conditions. Each point (specific activity values) was obtained from three different experiments and reported relative to wild type at 50°C (optimal conditions). Standard deviations do not exceed 5%. The wild type is shown as close circles and bold line. The curve for the quadruple mutant M4a is shown in orange and in red for M4c. The mutation R37A enhances the activity and shift the optimal temperature compared to wild type, the relative activity at 25 versus 60°C are listed in Table 1.

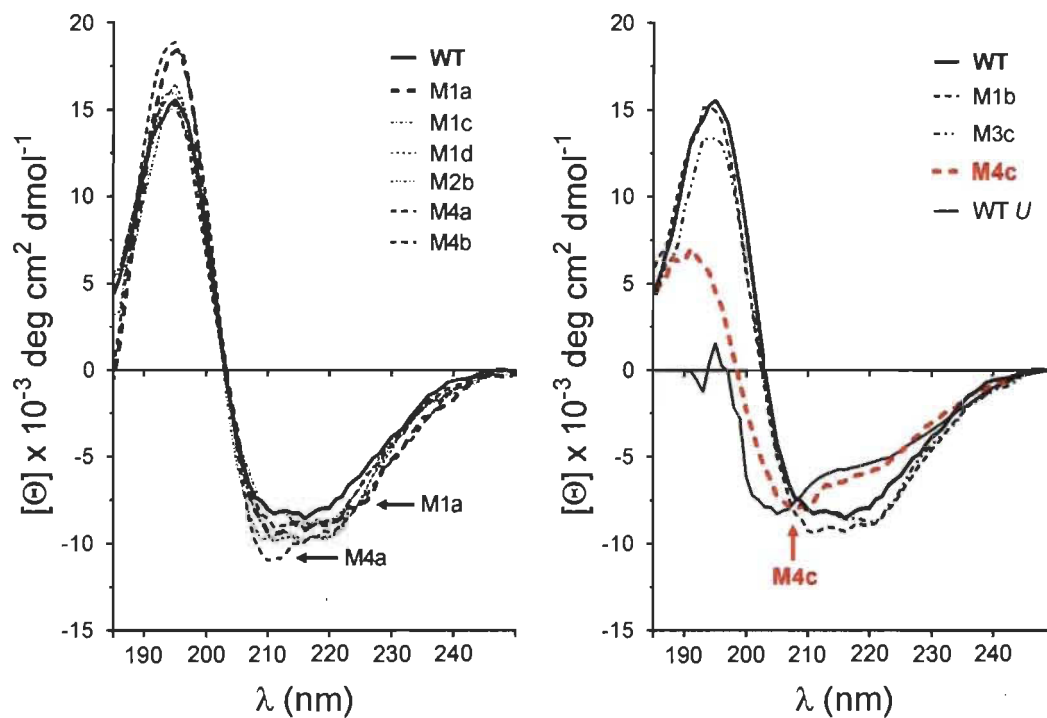


Figure 7. Conformational analysis by far-UV circular dichroism. Far-UV CD spectra were recorded at 25°C with 0.5 mg/mL of protein in 20 mM sodium phosphate pH 8. The wild type is shown in thick black lines. Left panel shows that most mutants are folded. Right panel shows misfolding of M4c. The quadruple mutant M4c spectra (red curves) were similar to spectra of unfolded EstGtA2 (WT *U*), wild type spectra recorded at 90°C.

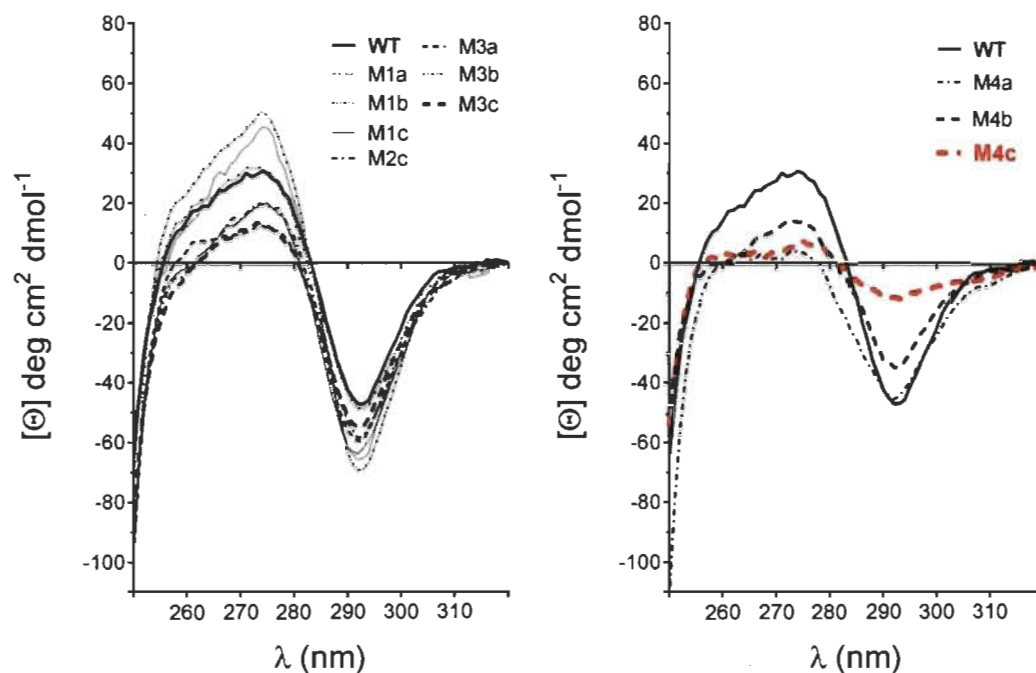


Figure 8. Conformational analysis by near-UV circular dichroism. Near-UV CD spectra were recorded at 25°C with 0.5 mg/mL of protein in 20 mM sodium phosphate pH 8. While all mutants had a strong minimum at 292 nm, M4c (red curve) had a weak signal throughout the range studied, suggesting loss of packing near aromatic residues.

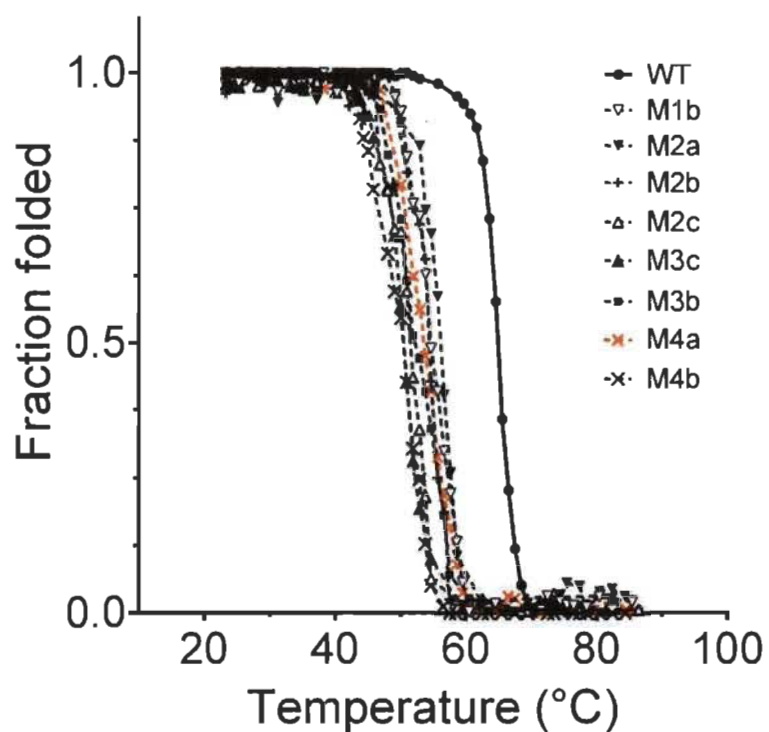


Figure 9. Thermal unfolding of EstGtA2 and mutants. Thermal unfolding curves recorded from 25 to 90°C were expressed as fraction folded derived from the CD signal at 222 nm as function of temperature. Samples were made of 0.5 mg/mL of protein in 20 mM sodium phosphate pH 8. EstGtA2 had the highest denaturation temperature among all versions studied. The quadruple mutant M4b is the least stable. The quadruple mutant M4a is shown in orange.

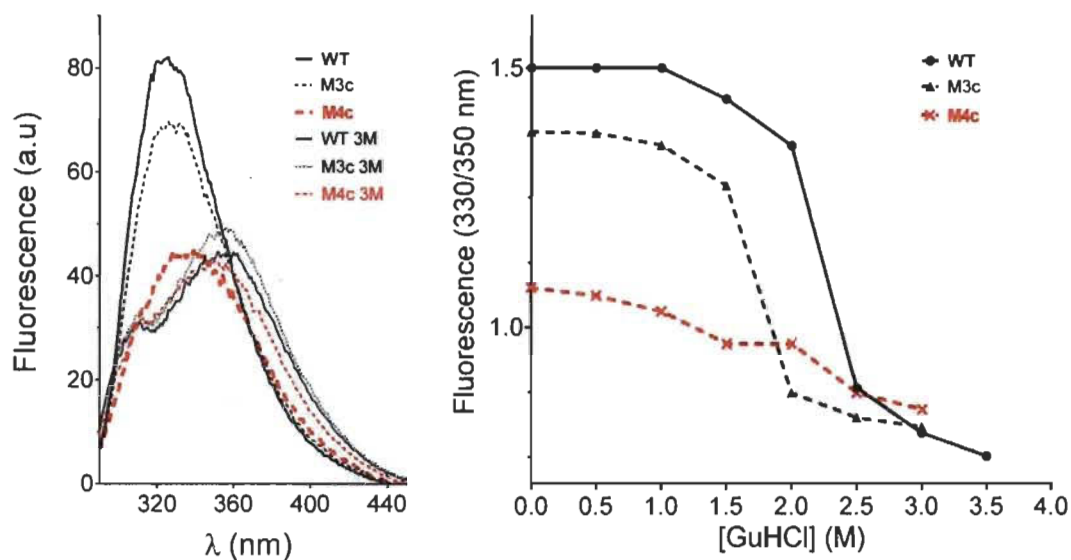


Figure 10. Intrinsic fluorescence. Panel A: Fluorescence spectra recorded for WT, M3c and M4c (red curves) at 25°C with 0.1 mg/mL at pH 8 in the same buffer (top spectra), and with 3 M GuHCl (bottom spectra). Panel B: Corresponding chemical denaturation curves are shown. Fluorescence peak was shifted toward longer wavelength upon denaturation, indicating hydration of aromatics. M4c showed hydration of aromatics in the absence of denaturant.

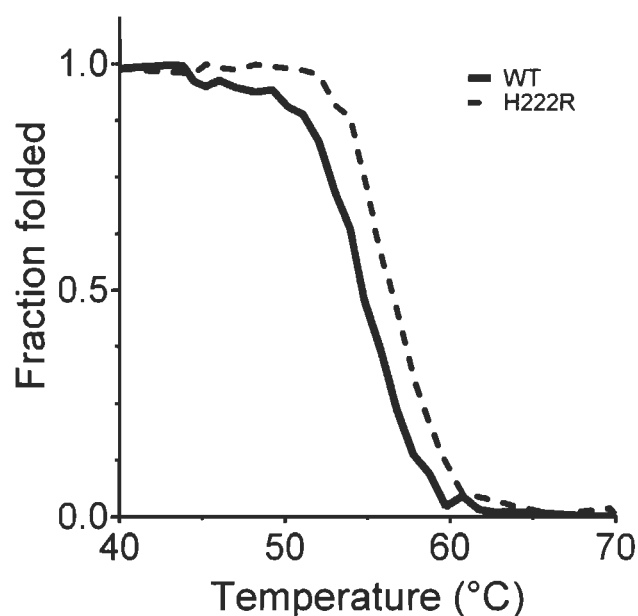


Figure 11. Modification of the interloop salt bridge in EstGtA2. The conserved inter-loop salt bridge found in EstGtA2 (D194-H222) was modified to resemble that of 3RM3 (D194-R222). Figure shows thermal unfolding curves calculated from the CD signal at 222 nm as function of temperature for the wild type enzyme and for the H222R mutant (protein concentration of 0.5 mg/mL in 20 mM CAPS pH 10). The mutation of H for R in position 222 led to increase the denaturation temperature by 2.1°C.

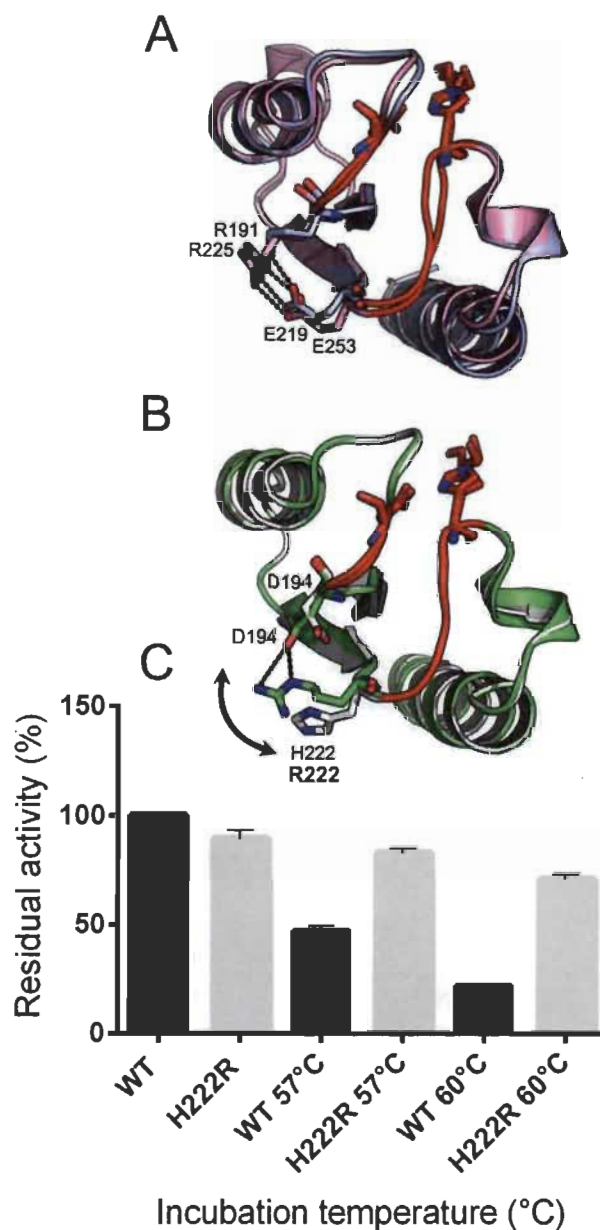


Figure 12. Impact of the new interloop salt bridge on EstGtA2 thermostability under alkaline pH. A. Structural alignment of the interloop portion for Est30 (blue) and LipS (pink). B. Structural alignment for MGL-H257 (green) and EstGtA2 (white), salt bridge at the conserved interloop position is shown for respective structures. Arrow shows change reversal for MGL-H257 and EstGtA2 (N' enzymes). The proteins (wild type or H222R mutant) were incubated for 10 min at the indicated temperature and the residual activity on *p*NP-octanoate (compared with non-incubated enzymes) was measured at 25°C in 20 mM CAPS buffer pH 10.

Supporting information

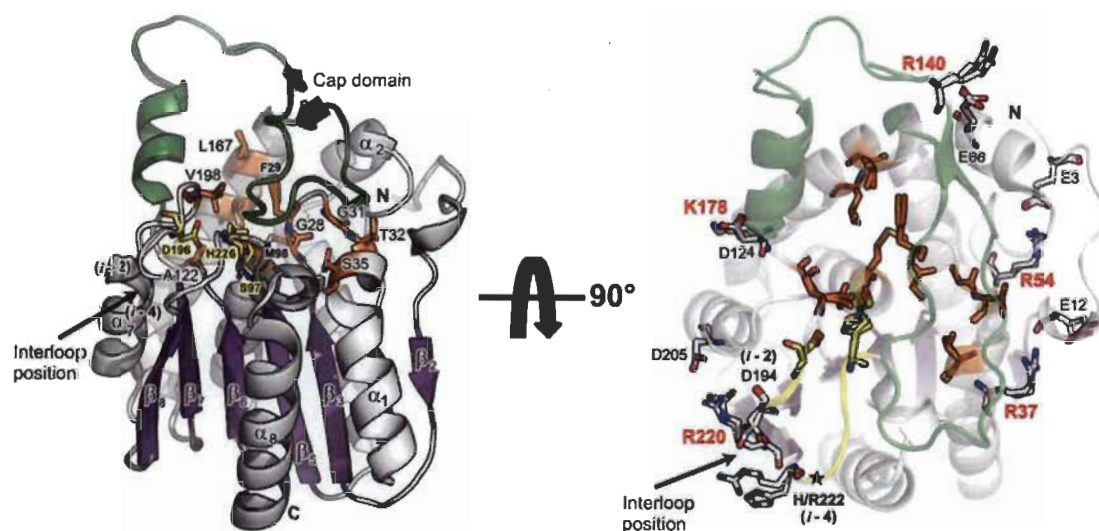


Figure S1. Structural model of EstGtA2. Structural alignment of EstGtA2 model and X-ray crystal structure of MGL H-257 (left) and with a 90° rotation (right). Conserved salt bridges studied are shown. Conserved residues of the binding site are shown in orange.

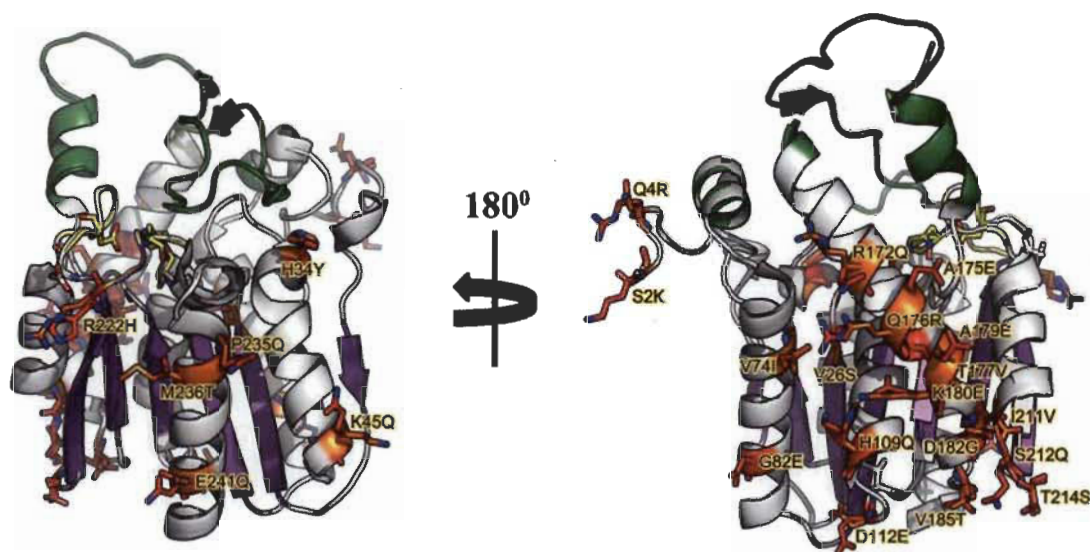


Figure S2. Amino acids substitution between EstGtA2 and MGL H-257. The 27 out of 249 residues that are different in EstGtA2 compared to MGL H-257 (89% identity) are shown in red (EstGtA2) and orange (MGL H-257).

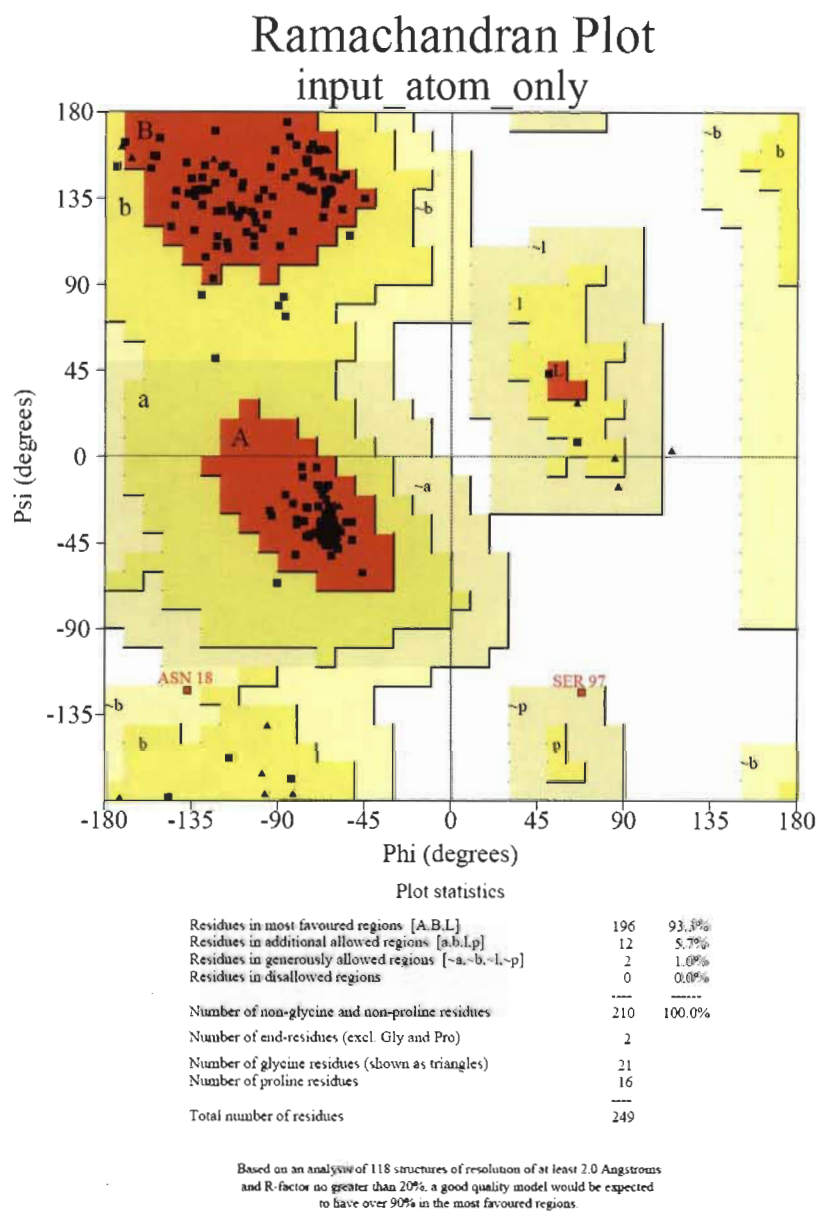


Figure S3. Refinement statistics for the EstGtA2 model. Ramachandran plot for the EstGtA2 model based on the X-ray crystal structure of MGL H-257 (PDB no. 3RM3).

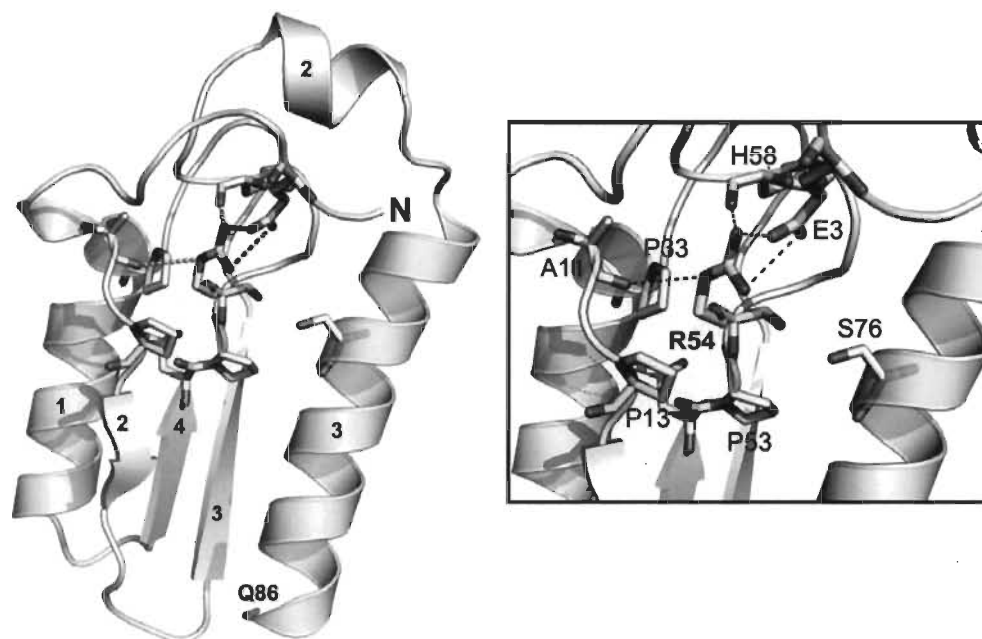


Figure S4. The E3-R54 salt bridge. The first 86 residues from the N-terminal end are shown. The E3-R54 salt bridge links the N-terminal end of EstGtA2 to the core. A hydrogen bond between the R54 guanidinium and the oxygen of H58 and A11 backbone is predicted. In addition the R54 side-chain would form a hydrophobic cluster with two prolines (P13 before strand β 2 and P33 after helix α 1).

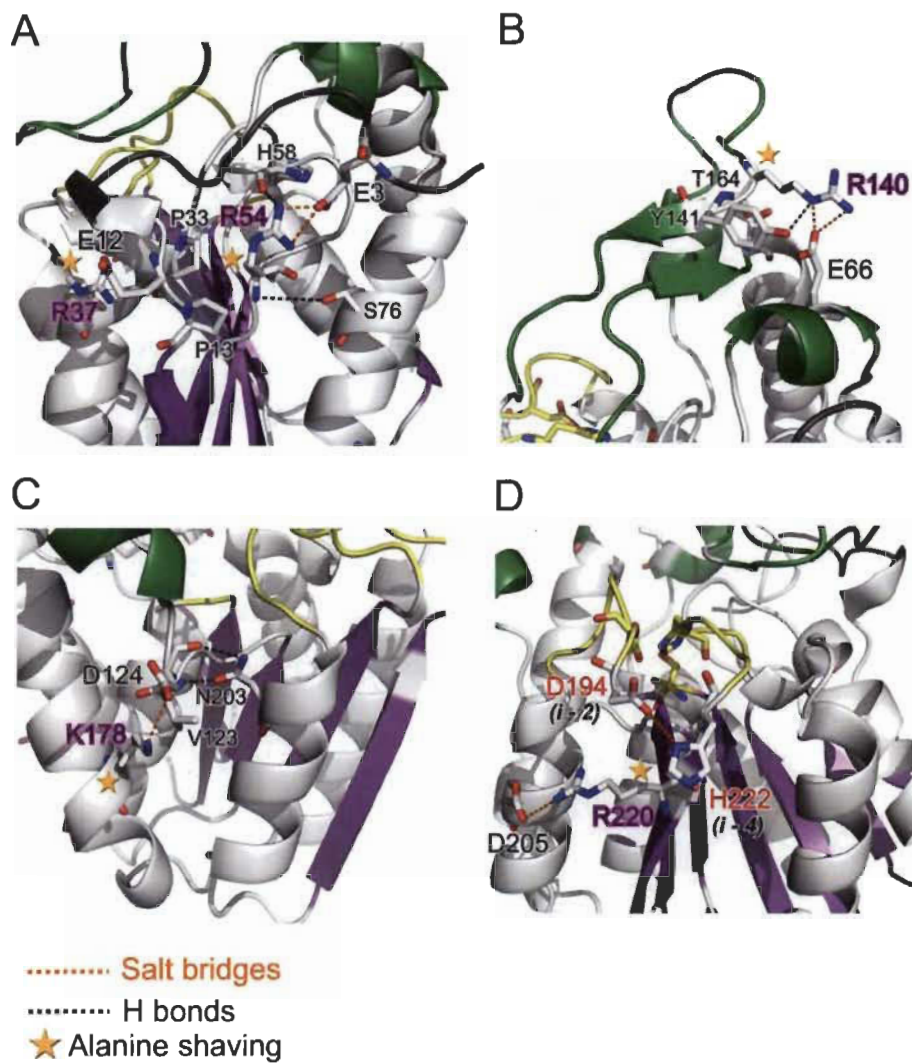


Figure S5. Salt bridges interactions. The predicted interactions for the following salt bridges: E3-R54 and E12-R37 (A), the E66-R140 (B), D124-K178 (C), D205-R220 and D194-H222 (D).

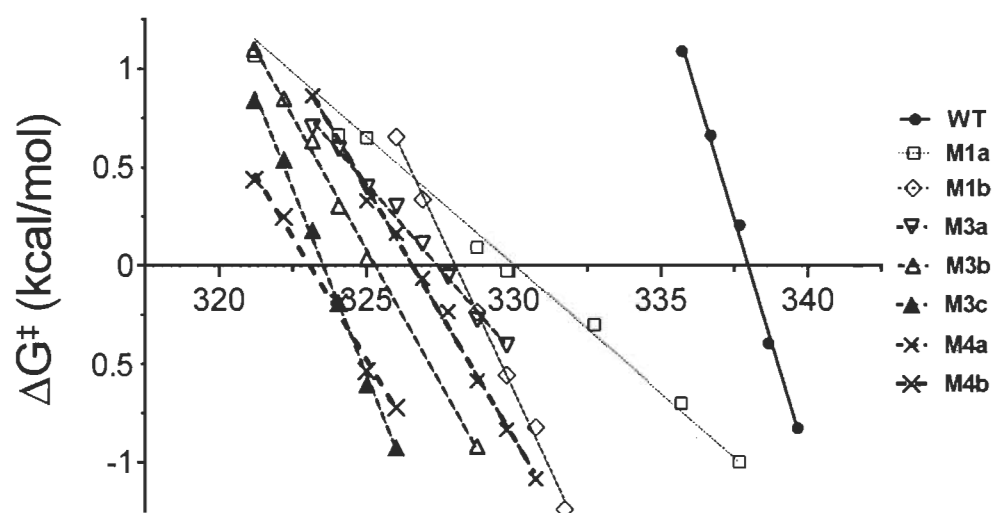


Figure S6. Unfolding free energy for EstGtA2 and mutants. The unfolding free energy (ΔG) as function of temperature (K) is shown. The melting temperatures (T_m) are determined at $\Delta G = 0$.

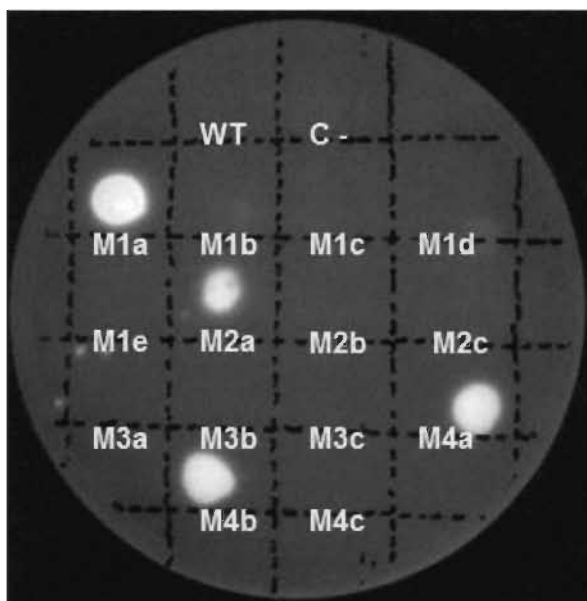


Figure S7. Hydrolysis of long-chain triglycerides by R37A. The wild type EstGtA2 and mutants were deposited (10 μ g) onto olive oil-agar plate containing rhodamine 0.001%. The hydrolysis of TAG released free fatty acids and the activity was detected under UV-illumination at 302 nm.

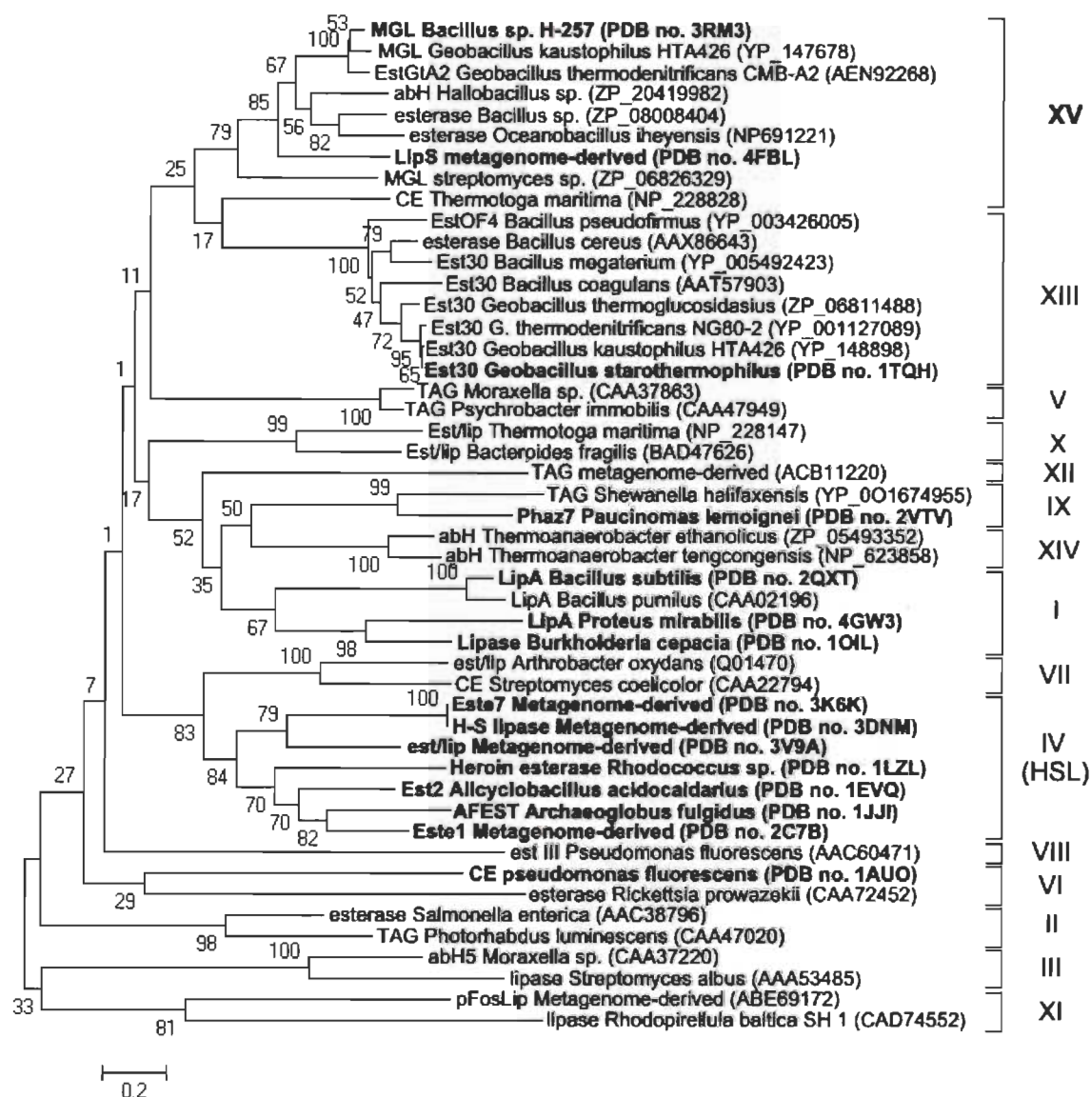


Figure S8. Phylogenetic tree showing the bacterial lipolytic enzymes classification. Protein sequences for which an X-ray crystal structure is solved are shown in bold within the tree. The novel cluster (family XV) is shown in bold right to the tree.

Table S1. Primers used for cloning and directed mutagenesis

Primers used for cloning		Sites / Vector	Reference
5' TAATTAGGTACCTATGAAAGAACGATATCCTGTACTT 3'	Fwd	<i>KpnI</i> / pQE31	Ref 7.
5' TATTAAGCTTTCAAGCATGTTTGGCGAA 3'	Rev	<i>HindIII</i> / pQE31	
5' CTCGGTACATATGAAAGAACGATATCCTGTACTTC 3'	Fwd	<i>NdeI</i> / pET28	This study
5' GTGCGGCCGTCAAGCATGTTTGGCGAAAACTG 3'	Rev	<i>EagI</i> / pET28	
Mutagenic primers		Mutation	Source
5' CACACAGCATGGCGCCGCTTGCTGAAGCGTATG 3'	Fwd	R37A	This study
5' CACACGCTTCAGCAAGCGGCGCCATGCTGTATG 3'	Rev		
5' GTTATACCGTTTGCTTGCCGGCGTTAAAGGGGCATGGGACGC 3'	Fwd	R54A	This study
5' GCGTCCCATGCCCTTTAACGCGCGGCAAGCAAACGGTATAAC 3'	Rev		
5' TGGCGAGGTGCCGGCGTATCTTGACTCGAT 3'	Fwd	R140A	This study
5' ATCGAGTCAAGATACGCCGGCACCTCGCCA 3'	Rev		
5' CTCAATTGATGGAGCGTGTGGCGGAAGAGCTTGGACGCATTAC 3'	Fwd	K178A	This study
5' GTAATGCGTCCAAGCTCTTCCGCCACACGCTCCATCAATTGAG 3'	Rev		
5' CAGAGAAAGAGATCGTCGCGCTTCACAACAGCTACC 3'	Fwd	R220A	This study
5' GGTAAGCTGTTGTGAAGCGCGACGATCTCTTTCTCTG 3'	Rev		
5' CTTTATTTGGCGGAACAGCGTCCGGAGATTTGCGGCATC 3'	Fwd	H110R	This study
5' GATGCCGCAAATCTCCGGACGCTGTTCCGCCAAATAAAG 3'	Rev		
5' GTTCCGATGAAGACCGTGTGTGCGCCTGG 3'	Fwd	H197R	This study
5' CCAGGCGGCACAACACGGTCTTCATCGGAAAC 3'	Rev		
5' GAGATCGTCCGTCTTCGTAACAGCTACCATGTAG 3'	Fwd	H222R	This study
5' CTACATGGTAGCTGTTACGAAGACGGACGATCTC 3'	Rev		

Table S2.**Table S2. Salt bridges studied.**

Acidic residues	Fraction accessible	Basic residues	Fraction accessible	Distance (Å)	Secondary structure
E3	0.18	R54	0.19	3.3	N-term/L5
E12	0.59	R37	0.20	3.9	L1/ α 1
E66	0.17	R140	0.74	3.6	α cap/Lcap
D124	0.40	K178	0.31	3.3	L10/ α 4
D205	0.59	R220	0.44	3.6	α 5/ β 8
D194	0.52	H/R222	0.90/0.88	4.0/3.9	L14/L16

Table S3.**Table S3. Distinctive salt bridges composition between the N', LipS and N cluster.**

N'	LipS	N
E3-R54 ^a	R66-E71	
E12-R37 ^a	E114-R142	
E66-R140 ^a	E115-R116	R37-E40
E78-H110	E141-K216	E124-K165
D124-K178 ^a	R154-E209	K139-E152
D148-H197 ^b	D179-H228 ^b	E142-K144
D194-H222	R225-E253	R191-E219
D205-R220 ^a	H233-E236	K216-D237
	E246-H279	
	E248-R271	

The conserved interloop salt bridge is shown in bold.

a. Salt bridges unique to N' family enzymes.

b. Conserved salt bridge between EstGtA2 and LipS

CHAPITRE VI

RÉSULTATS ET DISCUSSION

L'identification d'éléments structuraux au niveau de la séquence d'une enzyme pouvant permettre de prédire une propriété particulière comme sa thermostabilité permettrait d'améliorer l'efficacité de développement de nouveaux biocatalyseurs. L'évolution naturelle est le meilleur guide pour comprendre le rôle que peuvent jouer certains éléments structuraux au sein d'un groupe d'enzyme dans le contrôle d'une propriété comme la thermostabilité. La conservation de ces éléments structuraux au sein d'un groupe d'enzymes thermostables est souvent un bon indicateur de leur importance pour leur stabilité. En comparant la structure d'enzymes produites par les microorganismes thermophiles avec ceux de leurs contreparties mésophiles, différents motifs ont été identifiés comme responsables de la thermostabilité accrue des versions thermophiles. Un des éléments structuraux les mieux corrélés avec les protéines thermo- et hyperthermostables est la présence accrue de ponts salins. La coévolution de résidus chargés impliqués dans la formation de ponts salins au sein d'un groupe d'enzymes constitue un élément de structure identifiable au niveau de la séquence qui peut être utilisée pour l'annotation génomique comme marqueur séquentiel d'une propriété comme la thermostabilité.

Nous avons produit une nouvelle enzyme thermostable de *G. thermodenitrificans* appelé EstGtA2, modélisé sa structure et identifié une série de ponts salins conservés au sein d'un groupe particulier formant une nouvelle famille d'enzyme lipolytique. Nous proposons d'élargir la classification originale en ajoutant une 15^e famille d'enzymes lipolytiques bactériennes et suggérons d'utiliser la coévolution de ponts salins conservés au sein de la nouvelle famille XV et XIII comme marqueur séquentiel pour l'assignation de protéines à différentes sous-familles. Le rôle que jouent les ponts salins conservés, marqueur de la sous-famille N' dans la thermostabilité de EstGtA2 a été évalué de façon

combinatoire afin de déterminer l'importance de ces marqueurs structuraux sur la thermostabilité des ce groupe d'enzymes.

6.1 Approche choisie : mutagenèses rationnelles

Une approche classique par comparaison structurale et par mutagenèses dirigées a été mise en place pour identifier et comprendre le rôle de certains éléments structuraux impliqués dans la thermostabilité de EstGtA2 et d'utiliser cette compréhension pour améliorer sa stabilité en milieu alcalin. La structure de la nouvelle enzyme a été modélisée et les résidus chargés impliqués dans la formation de ponts salins ont été remplacés par des résidus alanine « alanine-scanning » afin de vérifier leur importance. D'autres mutations dirigées au niveau de résidus basiques ont été introduites dans le but de modifier la réponse au pH de la nouvelle enzyme lipolytique.

6.2 EstGtA2 : une nouvelle enzyme lipolytique thermostable

Une communauté de bactéries thermophiles se développant dans un compost organique à des températures entre 60 et 70 °C a été isolée. Le genre et l'espèce de chaque isolat ont été identifiés sur la base du gène d'ARNr 16S et par une caractérisation phénotypique et métabolique. Le profil d'activités hydrolytiques thermostables de chaque espèce a été évalué afin de cribler et d'entifier des producteurs d'enzymes lipolytiques. Deux des espèces identifiées ont montré un profil de production d'enzymes lipolytiques thermostables intéressants, soit *Geobacillus thermodenitrificans* et *Aneurinibacillus thermoaerophilus*. L'analyse du génome de la souche *G. thermodenitrificans* NG80-2 (NC_009328), isolée dans un puits de pétrole du nord de la Chine, a permis d'identifier environ six gènes annotés estérase/lipase putatifs et de synthétiser des amorces pour l'amplification d'un ORF particulier annoté estérase/lipase thermostable (YP_001125819.1). Les propriétés de cette nouvelle enzyme étaient encore inconnues et nous avons entrepris de caractériser l'enzyme homologue de (YP_001125819.1), que nous avons appelée EstGtA2, présent chez la souche de *G. thermodenitrificans* CMB-A2 isolée dans du compost organique fait au Québec.

L'ORF de EstGtA2 a été amplifié par PCR à partir de l'ADN génomique de *G. thermodenitrificans* CMB-A2. La séquence de l'ORF de EstGtA2 a été déposée dans GenBank (No. JN031579.1) ainsi que la séquence de EstGtA2 correspondante (No. AEN92268.1) et du gène d'ARNr 16S de la souche *Geobacillus thermodenitrificans* CMB-A2 (No. GQ293454.1). L'ORF encodant EstGtA2 a été cloné, et la protéine recombinante surexprimée chez *E. coli* (avec une étiquette de purification His₆-tag fusionnée en N-terminal), purifiée par chromatographie d'affinité (IMAC) et caractérisée pour la première fois (Figures B.4-B.9). La nouvelle enzyme catalyse l'hydrolyse d'une variété de para-nitrophenyl esters de différentes longueurs de chaînes (C4-C16) et catalyse aussi l'hydrolyse de la tributyrine sur une large gamme de pH. Par contre EstGtA2 est incapable de catalyser l'hydrolyse de triglycérides à longues chaînes (> C4), une propriété propre aux vraies lipases (EC 3.1.1.3). Vu le haut niveau d'identité de séquence entre EstGtA2 et la monoacylglycerol lipase MGL de *Bacillus* sp. H-257, il est probable que EstGtA2 soit une monoacylglycerol lipase (EC 3.1.1.23) catalysant l'hydrolyse de monoglycérides à longues chaînes, en plus de l'activité carboxylesterase identifiée (EC 3.1.1.1). Il n'est pas exclu que EstGtA2 puisse aussi montrer d'autres types d'activités hydrolytiques. À des fins de simplification, pour EstGtA2, nous référerons à l'activité carboxylesterase (EC 3.1.1.1).

La nouvelle EstGtA2 est active de pH 6 à 11 sur des esters de différentes longueurs de chaînes acylées et montre une activité optimale à 50 °C et pH 8 sur le substrat *p*NP octanoate (C8). EstGtA2 montre une stabilité conformationnelle de pH 4 à 10 avec des valeurs de T_m allant de 54 à 69 °C dépendamment du pH (Figure B.13). Quoique l'enzyme conserve une conformation repliée jusqu'à pH 4, aucune activité n'est détectée en dessous de pH 6 probablement dû au pK_a de l'histidine catalytique H226 qui nécessite d'être déprotoné pour la catalyse. La nouvelle EstGtA2 montre aussi une importante thermostabilité à 25 °C avec un temps de demi-inactivation d'environ 3.5 heures d'incubation à 80 °C et pH 8. Nous avons montré que cette thermostabilité suite à une incubation au-delà du T_m vient de la renaturation de EstGtA2, cette réversibilité de dénaturation thermique étant réduite par la présence du peptide exogène fusionné en N-terminal utilisé pour la purification (His₆-tag). À la température optimale

d'activité (50 °C), EstGtA2 montre une activité optimale à pH 8-9. La nouvelle enzyme montre aussi un profil de stabilité thermique hautement dépendant du pH. Les résultats de dichroïsme circulaire (CD) et de dispersion de lumière dynamique (DLS) montrent un état natif similaire en conditions acides et alcalines (propriétés spectrales similaires). En milieu alcalin (pH 7 à 10), un état dénaturé soluble permet une réversibilité de dénaturation thermique alors qu'en milieu acide (pH 4 à 6), une agrégation irréversible se produit suite à la dénaturation thermique rendant EstGtA2 incapable de se renaturer. La déplétion de ces agrégats insolubles après chauffage en milieu acide se traduit par une chute de signal CD et de polydispersité alors qu'en milieu alcalin les propriétés spectrales initiales sont retrouvées (Figures 6.1 et B.14-B.15). Cette stabilité conformationnelle dépendante du pH suggère que des interactions électrostatiques entre certains groupements ionisables sont impliquées dans la stabilité de EstGtA2.

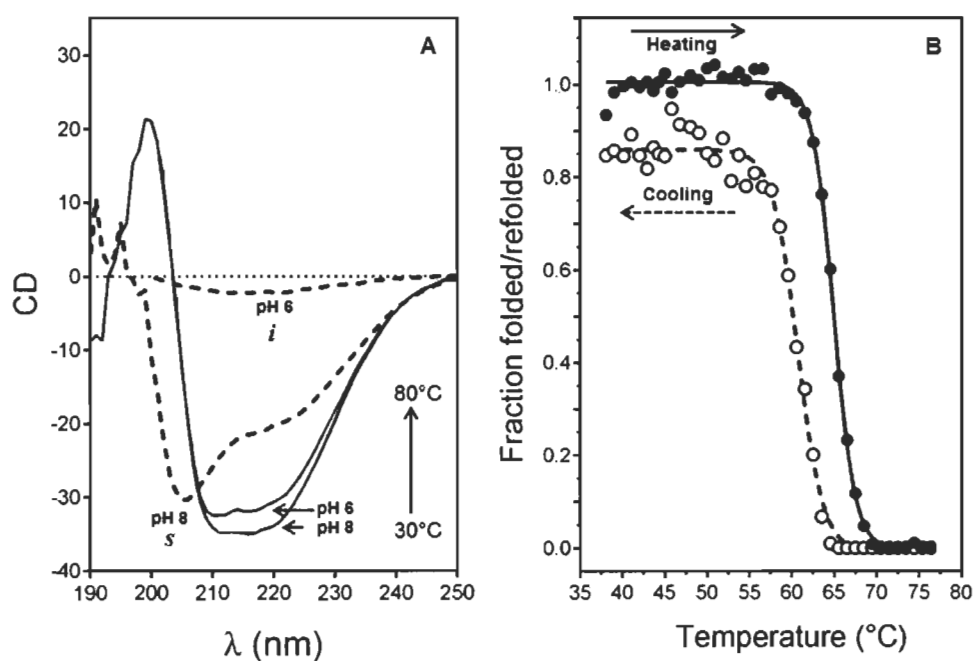


Figure 6.1 Réversibilité conformationnelle de EstGtA2.

A. Spectre de dichroïsme circulaire de EstGtA2 à pH 6 et 8 à 30 °C (lignes pleines) et 80 °C (pointillées). Un état dénaturé soluble à pH 8 (*s*) permet la renaturation alors qu'un état dénaturé insoluble (*i*) à pH 6 cause l'agrégation irréversible de EstGtA2. B. Courbe de dénaturation et de renaturation à pH 8 dérivée du signal CD à 222 nm mesurée de 30 à 80 °C à un taux de chauffage de 0.8°C/min et de 80 à 30 °C à un taux de refroidissement de 0.4°C/min (ligne pointillée).

6.3 Structure de EstGtA2

Initialement, la structure de EstGtA2 était inconnue et celle-ci a donc été modélisée sur la base de la structure cristalline de la carboxylesterase Est30 (32 % d'identité avec EstGtA2) de chez *G. stearothermophilus* (PDB no. 1TQH), résolue à 1.63 Å, le plus proche homologue pour laquelle une structure était élucidée et déposée dans PDB à l'automne 2009. Vu la faible identité de séquence entre les deux protéines, beaucoup d'efforts ont été déployés pour cristalliser EstGtA2 afin d'élucider sa structure par cristallographie aux rayons X. Une collaboration avec le professeur James Bowie à UCLA et ensuite avec le professeur Albert Berghuis à l'Université McGill ont été entreprises pour cristalliser EstGtA2, malheureusement aucun cristal d'une qualité adéquate n'a été obtenu, et ce, après le criblage de plusieurs milliers de conditions et avoir essayé de cristalliser quelques mutants. Un seul cristal a été obtenu au cours des criblages, mais ne diffractait pas les rayons X correctement vu sa forte hydratation. Une résolution de 2.9 Å et un patron de diffraction diffus nous ont obligés à utiliser ce cristal pour l'ensemencement de nouvelles conditions (Figures 6.2 et 6.3). Au printemps 2012, la structure de la monoacylglycerol lipase MGL de *Bacillus* sp H-257 (89 % d'identité avec EstGtA2) a été résolue à 1.2 Å et déposée dans PDB (no. 3RM3). La modélisation de EstGtA2 sur la base de cette nouvelle structure cristalline a permis d'obtenir un modèle d'une meilleure qualité (Figure 6.4). Le haut niveau d'identité de séquence entre ces deux enzymes (89 % d'identité) et les statistiques de raffinement de la structure suggère que le modèle de EstGtA2 généré est d'une qualité acceptable pour interpréter certains éléments structuraux responsables de lui conférer sa thermostabilité.

La comparaison du modèle de EstGtA2 avec la structure de MGL H-257 a révélé que les 27 résidus qui diffèrent entre les deux protéines (27 résidus sur 249 : 11 % de résidus non conservés) sont concentrés dans les hélices du corps alpha/bêta. Par ailleurs, la pochette de liaison ainsi que les résidus du site actif sont hautement conservés entre les deux protéines. L'analyse de la structure de MGL H-257 et du modèle de EstGtA2 a révélé une série de ponts salins conservés entre ces deux enzymes, lesquels n'avaient pas été observés dans le premier modèle initialement généré à partir de la structure de Est30 (PDB no. 1TQH).

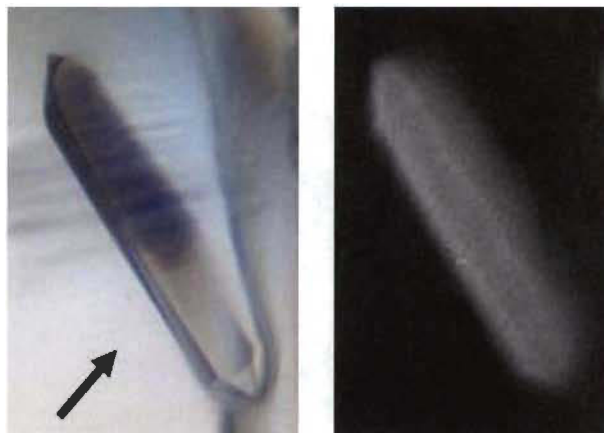


Figure 6.2 Cristal de EstGtA2.

Cristal obtenu lors des essais de cristallisation de EstGtA2. Conditions : 2M sulfate d'ammonium, 2 % 1,4 dioxane, 4 % PEG 400, 0.1 M HEPES, pH 7.5 à 25 °C. L'image de gauche montre une photographie microscopique en éclairage normal alors que l'image de droite montre le même cristal sous illumination UV.

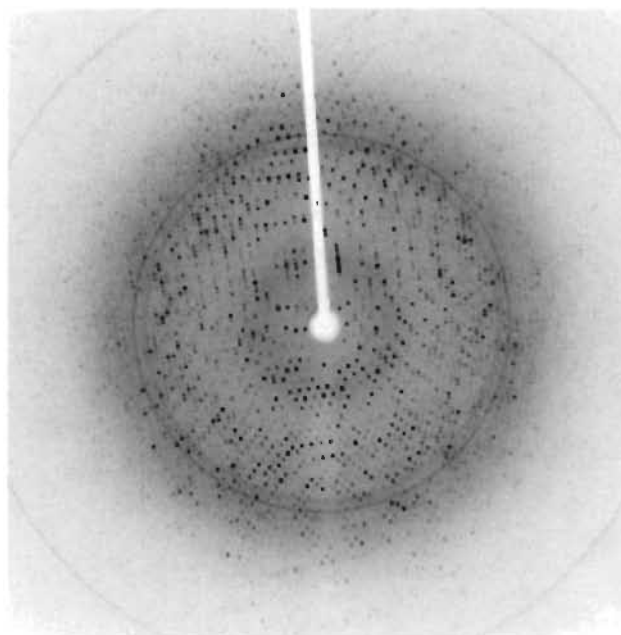


Figure 6.3 Photographie du patron de diffraction des rayons X de EstGtA2.

Les rayons X ont été générés avec une anode rotative Rigaku et le patron de diffraction collecté à l'aide d'un détecteur Raxis IV ++. Le patron révèle un cristal fortement hydraté et une résolution de 2.9 Å. Le patron a été enregistré en mai 2012 à l'Université McGill.

Une comparaison structurale entre EstGtA2 et MGL H-257 a permis d'identifier 8 ponts salins conservés, 3 impliquant des résidus histidine et 5 impliquant des résidus arginine ou lysine (Figures 6.4 et B.16). Des alignements multiples de séquences ont montré que les résidus aux positions correspondantes à ces ponts salins sont très bien conservés chez d'autres espèces thermophiles ce qui suggère que ceux-ci jouent un rôle important dans la stabilité de ces enzymes. Le haut niveau de conservation de ces ponts salins ainsi que les conditions de cristallisation (pH 6.5-7) suggère que ces ponts salins sont formés dans EstGtA2 en solution à pH 8 (région étudiée). De plus, les enzymes homologues présentant une conservation des résidus impliqués dans la formation de ces ponts salins fonctionnent optimalement entre pH 8 et 9.5.

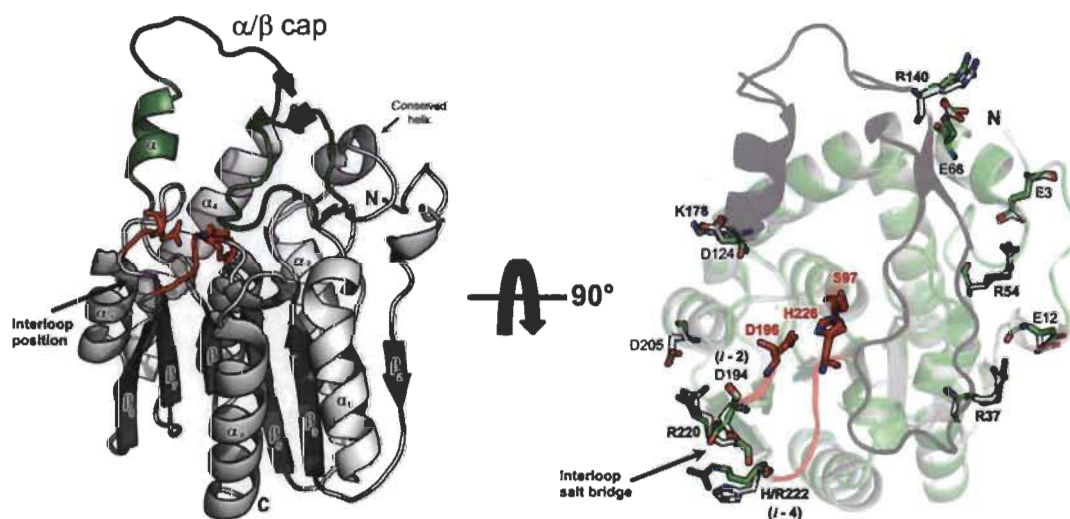


Figure 6.4 Structure tridimensionnelle de EstGtA2.

Modèle de EstGtA2 (à gauche) et alignement structural avec la structure cristalline de MGL H-257 (PDB no. 3RM3). La triade catalytique et les différents ponts salins conservés sont montrés.

6.3.1 Une 15^e famille d'enzymes lipolytiques bactériennes (famille XV)

Les enzymes lipolytiques bactériennes ont été initialement classées en huit familles distinctes sur la base de l'identité de séquence, de motifs conservés et des propriétés biochimiques (famille I-VIII) par Arpigny et Jaeger. Nous avons suggéré l'existence d'une nouvelle famille de carboxylesterase dans laquelle pourrait être classé EstGtA2 étant donné qu'aucun des critères établis pour assigner les enzymes

lipolytiques aux huit familles n'était rencontré pour EstGtA2. Une analyse phylogénétique a permis de comprendre que la séquence de EstGtA2 ainsi que celle des homologues montrant une conservation des résidus possiblement impliqués dans la formation de ponts salins forment un clade phylogénétique distinct. Ce groupe d'enzymes qui n'appartient à aucune des huit familles d'enzymes lipolytiques bactériennes proposées dans la classification originale a été appelé N'. Par la suite, six autres familles ont été ajoutées à la classification originale (famille IX-XIV) sur la base de groupe phylogénétique distinct. Le groupe d'enzymes dont fait partie EstGtA2, N', sort encore du lot du point de vue phylogénétique et ne présente aucun des motifs identifiés pour les autres familles. Nous proposons donc une 15^e famille d'enzymes lipolytiques bactériennes (famille XV) et suggérons des critères pour reconnaître les enzymes appartenant à cette nouvelle famille ainsi que pour l'assignation de nouvelles séquences à cette famille. Quoique la famille XV (EstGtA2) présente plusieurs similarités avec la famille XIII (Est30), suffisamment de divergence est observée entre ces deux groupes. Ces divergences sont à la base des critères établis pour la définition de la nouvelle famille XV.

Critère no 1. Les enzymes lipolytiques de la famille XV forment un groupe (clade) phylogénétique distinct de la famille XIII ou des autres familles.

Critère no 2. Les enzymes lipolytiques de la famille XV montrent une architecture du cap de type α/β contrairement à l'architecture dite α de la famille XIII.

6.3.2 Le groupe N', un sous-clade de la nouvelle famille XV

Similairement, nous avons établi des critères supplémentaires qui permettent de reconnaître rapidement les enzymes lipolytiques de la sous-famille N'.

Critère no 1. Les enzymes appartenant à la sous-famille N' présentent une polarité (- / +) au niveau du pont salin interboucle conservé en position ($i - 2$, $i - 4$) de l'aspartate et de l'histidine catalytique respectivement.

Critère no 2. Les enzymes de la sous-famille N' présentent un patron de ponts salins conservés. Cinq ponts salins sont exclusivement conservés chez les enzymes du clade phylogénétique correspondant à la sous-famille N'. La structure de ces ponts salins conservés est montrée à la Figure 6.5.

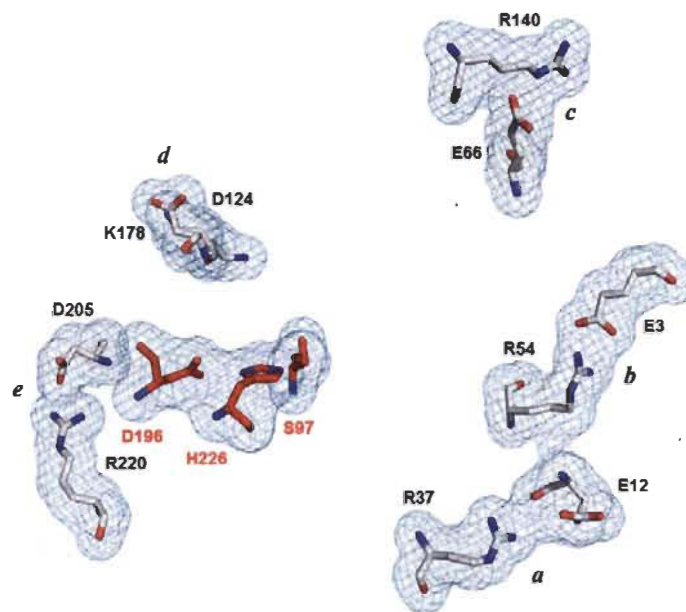


Figure 6.5 Structure des ponts salins étudiés dans EstGtA2.

Densité électronique ($F_o - F_c$) et représentation en bâtonnet des résidus impliqués dans la formation des ponts salins conservés dans la famille N' : E12-R37, E3-R54, E66-R140, D124-K178 et D205-R220 (a-e). Les atomes de carbone sont représentés en gris alors que les atomes d'oxygène et d'azote sont représentés en rouge et en bleu respectivement. La triade catalytique (S97, D196 et H226) est montrée en rouge (atomes de carbone et d'oxygène).

6.3.3 Rôle des ponts salins conservés dans la thermostabilité de EstGtA2

Le rôle que jouent les ponts salins conservés au sein des enzymes de la famille N' dans la stabilité de EstGtA2 a été évalué par une approche de mutagenèse combinatoire de type « alanine-scanning ». Cette méthode consiste à remplacer un résidu d'acide aminé susceptible d'être impliqué dans la propriété que l'on veut étudier par un résidu alanine et de vérifier l'impact de cette substitution sur la propriété en question. Cette approche est largement utilisée pour déterminer les résidus impliqués dans la liaison de

substrats au niveau du site actif d'une enzyme ou encore d'un ligand à son récepteur par exemple. La méthode normalement utilisée pour déterminer la contribution d'un pont salin à la stabilité d'une protéine est la méthode du cycle à double mutant (DMC). Cette méthode consiste à substituer chacun des résidus formant ce pont, individuellement, et en même temps, par un résidu isostère ($D \rightarrow N$ et $E \rightarrow Q$) et de comparer l'énergie libre de dénaturation entre les versions et avec le type sauvage. On compare la version originale (le pont salin formé), la version où la charge négative du résidu acide est annihilée, la version où c'est plutôt la charge positive qui est retirée et finalement la version double mutant (où les deux charges sont enlevées en même temps). Cette approche permet d'estimer l'énergie dite de couplage associée à un pont salin, mais ne permet pas d'évaluer la contribution de plus d'un pont à la fois. Nous avons donc choisi la méthode par « alanine-scanning » puisque nous voulions comprendre le rôle que jouaient ces ponts salins multiples de façon combinatoire.

De plus, nous avons choisi de substituer les résidus basiques, chargés positivement pour cette étude. Les résidus arginine ou lysine ont été ciblés, les résidus histidine n'ont pas été considérés étant donné leur faible pK_a (~ 6.1) et la faible proportion de ponts pouvant être formés dans la plage de pH alcaline étudiée (pH 8-10). Cinq ponts salins conservés ont été rompus de façon combinatoire par cette approche. Il est important de noter que le remplacement de résidus volumineux tels que l'arginine ou la lysine pour une alanine équivaut littéralement à raser ou couper la chaîne latérale en plus d'enlever la charge. Le pont hydrogène, l'interaction électrostatique ainsi que toutes autres interactions venant de la chaîne latérale des résidus arginine et lysine sont annihilées par la substitution pour une alanine. Par contre, la substitution pour une alanine permet aussi de réduire le nombre de nouvelles interactions pouvant se former, contrairement à d'autres types de substitutions pour lesquelles des effets collatéraux indésirables peuvent survenir. À titre d'exemple, l'équipe de Brian Matthews a évalué l'impact de la substitution d'un résidu arginine impliqué dans la formation d'un réseau de ponts salin par chacun des 19 autres acides aminés dans la stabilité du lysozyme du phage T4. Les résultats démontrent que le remplacement d'un résidu arginine par une alanine est beaucoup moins déstabilisant que la plupart des autres substitutions. En fait, le

remplacement de R par A est la troisième mutation la moins déstabilisante après le remplacement $R \rightarrow K$ et $R \rightarrow Q$ (Mooers *et al.*, 2009; Walter *et al.*, 1995). Par contre, le remplacement $R \rightarrow Q$ a dans certains cas des effets collatéraux indésirables (Stern and Khorana, 1989; Takeda *et al.*, 1990). Nous avons donc choisi de remplacer les résidus arginine pour des résidus alanine afin d'étudier l'impact global et l'importance totale que peuvent jouer ces résidus chargés impliqués dans la formation de pont salin dans la thermostabilité de EstGtA2 (Figures 6.6 et B.17). Les ponts salins suivants ont été étudiés : E3-R54, E12-R37, E66-R140, D124-K178 et D205-R220. Chacun des cinq résidus basiques impliqués (soit R37, R54, R140, K178 et R220) a été substitué par un résidu alanine (mutation $R/K \rightarrow A$). Les résultats montrent que lorsqu'un de ces ponts est brisé (une chaîne latérale basique « rasée »), la stabilité conformationnelle de EstGtA2 est considérablement réduite.

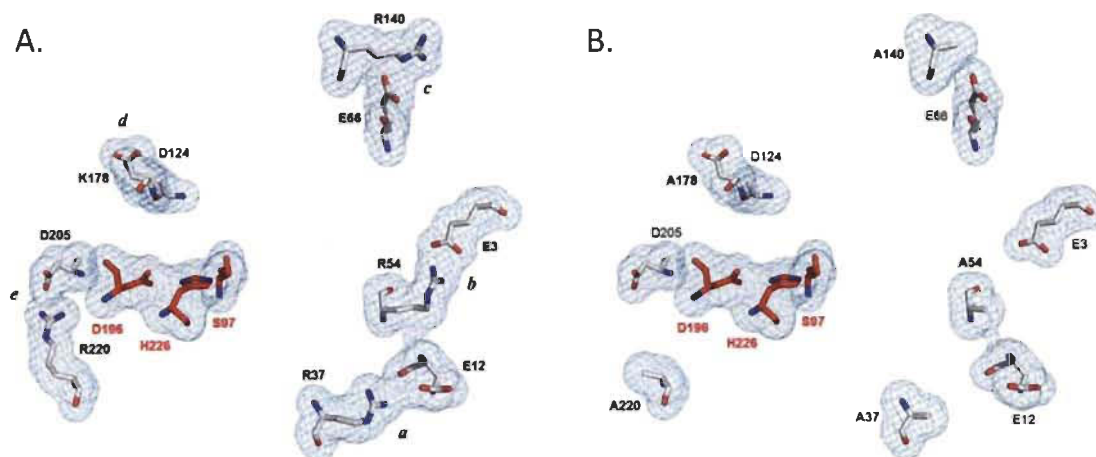


Figure 6.6 Localisation des substitutions alanine multiples.

A. Densité électronique (Fo-Fc) et représentation en bâtonnet des résidus impliqués dans la formation des ponts salins étudiés, annotés dans l'ordre séquentiel des résidus chargés positivement (a-e). Les résidus arginine et lysine (R37, R54, R140, K178 et R220) ont été substitués de façon combinatoire par un résidu alanine (A). La triade catalytique (S97, D196 et H226) est montrée en rouge. B. Localisation des cinq substitutions alanine ($R/K \rightarrow A$). Cinq simples mutants : R37A (M1a), R54A (M1b), R140A (M1c), K178A (M1d) et R220A (M1e), ainsi que trois doubles (M2a-c), trois triples (M3a-c) et trois quadruples mutants (M4a-c) ont été étudiés.

Afin d'évaluer l'impact de ces substitutions sur l'activité de EstGtA2, l'hydrolyse du *p*NP octanoate catalysée par EstGtA2 et les différents mutants a été mesurée sous la saturation, une condition qui reflète l'efficacité de l'enzyme (k_{cat}/K_M) en fonction de la température. Les simples substitutions entraînent une légère augmentation de l'activité à basse température, possiblement due à un accroissement de la flexibilité structurale. Par contre, lorsque deux résidus basiques ou plus sont mutés en même temps, l'activité de EstGtA2 est considérablement réduite à basse température, celle-ci devenant pratiquement nulle à haute température. Les différentes combinaisons de mutations ont un effet négatif sur la thermostabilité de EstGtA2, avec des diminutions de T_m allant de -5 à -14 °C et des impacts sur l'énergie libre de dépliement allant de $\Delta\Delta G_{25\text{ °C}} = -0.9$ à -6.9 kcal/mol pour les simples et les quadruples mutants respectivement (Figures 6.7 et B.18-B.19).

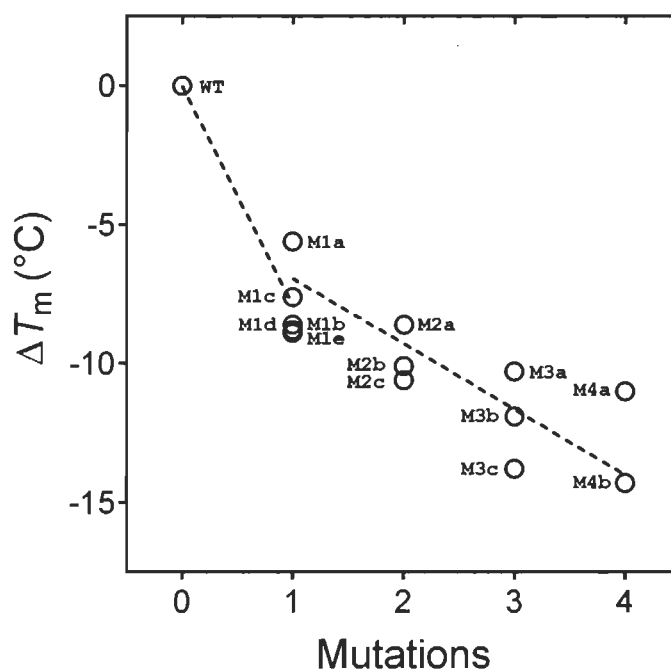


Figure 6.7 Impact des substitutions sur la stabilité EstGtA2.

La différence de température de fusion $\Delta T_m = T_m(\text{mutant}) - T_m(\text{WT})$ est représentée pour chaque mutant (les simples mutations M1a-e, doubles mutants M2a-c, triples mutants M3a-c et les quadruples mutants M4a-b). La relation montre l'impact associé à l'accumulation des mutations (substitutions alanine) sur la stabilité conformationnelle de EstGtA2 mesuré par dichroïsme circulaire à pH 8.

Une chute d'activité enzymatique à 60 °C allant de -8 % jusqu'à une complète inactivation est observée pour certains triples et quadruples mutants démontrant que ces résidus chargés impliqués dans la formation de ponts salins sont essentiels pour l'activité à haute température de EstGtA2 et pour sa thermostabilité. Une étude similaire réalisée par Pezzulo *et al.*, dans laquelle les auteurs ont analysé l'impact de la substitution de résidus chargés impliqués dans la formation de ponts salins exposés en surface pour des résidus alanine (alanine-scanning) a démontré que les substitutions étaient déstabilisantes avec des impacts négatifs sur le T_m allant de 1 à 15 °C, ces impacts étant davantage drastiques pour les substitutions de résidus arginine (Pezzullo *et al.*, 2012).

6.3.4 Rôle des ponts salins conservés dans le repliement de EstGtA2

Nous avons démontré que certains résidus impliqués dans la formation de ponts salins sont importants, voire essentiels pour la thermostabilité de EstGtA2. Quoique EstGtA2 montre une haute tolérance aux substitutions non conservatives, pouvant tolérer jusqu'à 4 substitutions (R/K→A), adoptant une structure proche de celle native (propriétés spectrales similaires en CD), une certaine combinaison a excédé cette tolérance. Le quadruple mutant M4c contient une combinaison de mutations qui empêche ou qui déstabilise EstGtA2 à un tel niveau la rendant incapable de se replier causant son inactivation complète à 25°C. Le mécanisme par lequel ces quatre résidus basiques (R54, R140, K178 et R220) qui forment des ponts salins semi-enfouis avec (E3, E66, D124 et D205) sont impliqués dans le repliement de EstGtA2 reste obscur, mais les résultats suggèrent que les interactions formées par ces résidus, soit les interactions spécifiques par la formation de ponts salins ou les effets de répulsion peuvent jouer un rôle dans le repliement de EstGtA2. Est-ce la perte de la contribution électrostatique, du pontage hydrogène, des interactions venant de la chaîne latérale aliphatique de ces résidus basiques (R et K), ou simplement la création de cavités pour certains résidus semi-enfouis ou bien une combinaison de ces facteurs qui est responsable de la déstabilisation? Différentes études ont démontré que la contribution moyenne d'un pont salin enfoui ou semi enfoui sur la stabilité d'une protéine varie de 3-5 kcal/mol (Waldburger *et al.*, 1995; Tissot *et al.*, 1996; Vijayakumar *et al.*, 2001).

L'énergie libre de stabilisation associée au repliement de EstGtA2, estimée à partir des états de transitions en CD varie de $\Delta G_{25^\circ\text{C}} = 8\text{-}17$ kcal/mol en fonction du pH, ce qui s'accorde avec les valeurs rapportées dans la littérature pour des enzymes monomériques thermo- et hyperthermostable de la taille de EstGtA2, soit de 5-30 kcal/mol (Backmann *et al.*, 1998; Wasenberg *et al.*, 1999; Deutschman *et al.*, 2001; Luke *et al.*, 2007).

Dans la région de pH étudiée (pH 8), EstGtA2 montre un $\Delta G_{25^\circ\text{C}}$ moyen de ~ 12 kcal/mol. Des contributions importantes à l'énergie de stabilisation sont observées pour les différents résidus de ponts salins. Des impacts sur l'énergie libre de stabilisation ($\Delta\Delta G_{25^\circ\text{C}}$) de -0.9 à -3.1 kcal/mol sont observés pour les différentes combinaisons de substitutions simples et doubles, alors que l'impact venant de la combinaison des différentes substitutions de trois et quatre résidus basiques varie de $\Delta\Delta G_{25^\circ\text{C}} = -3.6$ jusqu'à -6.9 kcal/mol. Le mutant M4a montre la plus importante déstabilisation ($\Delta\Delta G_{25^\circ\text{C}} = -6.9$ kcal/mol) avec un $\Delta G_{25^\circ\text{C}} = 4.9$ kcal/mol, ce qui est près des valeurs limites observées pour le repliement et la stabilité conformationnelle de ce type d'enzyme, mais adopte une conformation proche de celle native. Il se peut que pour le quadruple mutant M4c, lequel est incapable d'adopter la conformation native, l'accumulation des mutations et l'impact sur la stabilité aient excédé l'énergie libre associée au repliement de EstGtA2. Il est important de noter que jusqu'à 50 % d'identité de séquence est observé au sein des enzymes de la famille XV qui adoptent une structure très similaire alors que le mutant M4c (R54A/R140A/K178A/R220A) ayant 98 % d'identité de séquence avec EstGtA2 est incapable de se replier. Les éléments spécifiques d'un repliement particulier se trouvent-ils dans le profil d'hydrophobicité/hydrophilicité du repliement et/ou seraient-ils localisés au niveau de la séquence?

Quel est le rôle des répulsions électrostatiques dans le repliement des protéines? Les résidus impliqués dans la formation de ponts salins étudiés dans EstGtA2 sont clairement importants pour sa thermostabilité. Ceux-ci étant conservés dans les enzymes de la sous-famille N', ont-ils été sélectionnés pour leur rôle stabilisant dans la structure

native ou ont-ils quelque chose à voir dans le repliement, que ce soit par effet de répulsion avec d'autres résidus chargés ou par la formation de ponts salins transitoires?

Robert Sauer et son équipe ont étudié l'impact de différentes substitutions sur la stabilité et le repliement du répresseur Arc P22. Une étude particulièrement intéressante explore la mesure par laquelle le répresseur Arc P22 peut tolérer des substitutions de résidus alanine. Jusqu'à 15 substitutions (X→Ala) peut être toléré par le répresseur Arc P22 (dans la même protéine) sans changer sa conformation et entraîner d'impact négatif sur le T_m ou les propriétés de liaison à l'ADN. Le récepteur Arc P22 est une petite protéine de 53 acides aminés, c'est donc 28 % des résidus (au sein de la même protéine) qui peuvent tolérer une substitution pour une alanine sans effets négatifs sur ces propriétés. Il n'est pas surprenant de constater que la majorité de ces substitutions neutres soient observées à la surface de la protéine, mais curieusement, certains résidus enfouis formant le cœur hydrophobe de Arc P22 ainsi que d'autres résidus impliqués dans la formation de ponts hydrogènes et de ponts salins peuvent être substitué par un résidu alanine sans affecter les propriétés de Arc P22. Par contre, environ 50 % des positions de la séquence de Arc P22 ne tolère pas une substitution par un résidu alanine. Ces substitutions affectent la majorité des résidus du cœur hydrophobes, des résidus formant des ponts hydrogènes et des ponts salins. Les mutations les plus déstabilisantes concernent les résidus non polaires enfouis au centre du cœur hydrophobe. Un pont salin enfoui dans le cœur hydrophobe semble très stabilisant. Le remplacement des résidus de ce pont salin pour des résidus alanine déstabilise Arc P22 par 3-6 kcal/mol-dimère (Marqusee et Sauer, 1994; Sauer *et al.*, 1996; Brown et Sauer, 1999). Ces résultats suggèrent fortement que ce soit un nombre limité des résidus de la séquence d'une protéine qui détermine son repliement et sa stabilité. S'il existe un code pour le repliement des protéines en solution, il n'est pas impossible que celui-ci implique certains résidus ou segments très discrets au niveau de la séquence d'une protéine.

Nos résultats démontrent l'importance de ces larges résidus chargés (arginines et lysines) sur le repliement et la thermostabilité de EstGtA2. Le fait que ces résidus soient impliqués dans la formation de ponts salins relativement enfouis peut expliquer

l'importance capitale que jouent ces résidus dans le repliement et la stabilité de EstGtA2. Historiquement, les ponts salins étaient perçus comme la force dominant le repliement des protéines. Différentes études ont par la suite démontré la dominance des effets hydrophobes dans le processus du repliement. Par contre, il n'est pas exclu que dans certains cas et dans certaines conditions, que les interactions électrostatiques puissent jouer un rôle dans le repliement des protéines, soit par la formation de ponts salins ou par des effets de répulsion. Roca *et al.*, ont démontré que les interactions électrostatiques et les ponts salins peuvent être traités comme d'importants déterminants de l'énergie pour le repliement de certaines protéines (Roca *et al.*, 2007). Azia et Levy ont par ailleurs démontré que les interactions électrostatiques peuvent moduler le repliement de NTL9 et RNase (Azia et Levy 2009). Le fait que les ponts salins jouent un rôle important dans la cinétique du repliement de la Barnase avait aussi déjà été rapporté (Tissot *et al.*, 1996).

6.3.5 Modification d'un pont salin clé en position ($i - 2$, $i - 4$)

Parmi les huit ponts salins identifiés dans la structure de EstGtA2, un en particulier montre une caractéristique intéressante. Le pont salin entre D194 et H222 dans EstGtA2 est trouvé à une position particulière : il est exposé près du site actif et relie deux boucles catalytiques. Ce pont est formé, entre un résidu acide (D194) situé en position $i - 2$ de l'aspartate catalytique et un résidu basique (H222) situé en position $i - 4$ de l'histidine catalytique. Une analyse par alignement multiple de séquences montre une particularité intéressante associée à ce pont. Un haut niveau de conservation de résidus de charges opposées a été détecté à cette position, toujours situés en $i - 2$ et $i - 4$ des résidus catalytiques, formant possiblement toujours un pont salin, mais avec une haute tolérance aux mutations. Toutes les possibilités de partenaires chargés ont été identifiées. Par exemple, le pont D194-R222 est trouvé dans la structure cristalline de MGL H-257 (3RM3) alors qu'à cette même position c'est un pont formé par un résidu arginine et un résidu glutamate, en inversion de polarité d'interaction qui est trouvée dans la structure cristalline de LipS (4FBL) ainsi que celle de Est30 (1TQH). Vu la haute conservation d'un pont salin à cette position et la tolérance aux mutations, nous avons choisi

d'améliorer de façon rationnelle la thermostabilité en milieu alcalin de EstGtA2 en modifiant ce pont. Étant donné que EstGtA2 montre une activité et une stabilité optimale en milieu alcalin (pH 8-10), et vu le faible pKa de l'histidine (H222) exposée en position $i - 4$ de de l'histidine catalytique (H226), la substitution rationnelle par une arginine devrait permettre de former un pont salin dans la plage de pH alcalin étudiée vu sa présence dans la structure cristalline de MGL H-257 et la haute tolérance aux mutations observée. Comme prévu, la substitution H222R a augmenté la température de fusion de EstGtA2 de 2.1 °C à pH 10 et d'environ 3 fois sa thermostabilité enzymatique. Cet accroissement de la stabilité est observé de façon maximale à pH 10. Le succès de cette mutation rationnelle pour la stabilisation de EstGtA2 est possiblement dû à la position interboucle particulière. Le haut niveau de conservation d'un pont salin et la haute tolérance aux mutations et aux inversions de polarité de charges à cette position expliquent probablement la stabilisation obtenue par cette substitution rationnelle.

6.3.6 Modifications d'autres ponts salins dans EstGtA2

La stabilisation effective par la substitution H222R et la formation probable du pont salin interboucle D194-R222 à pH alcalin nous a poussés à évaluer l'impact d'une substitution H→R aux deux autres positions où le pont salin implique un résidu histidine, les ponts E78-H110 et D148-H197, (ces résidus histidine ayant été omis lors de l'étude par « alanine-scanning » vue leur faible pKa). Les mutations H110R et H197R ont été introduites individuellement et de façon itérative dans le mutant H222R dans le but d'accroître la stabilité de EstGtA2 en milieu alcalin. Les résultats montrent que ces doubles mutants ont chacun une température de fusion réduite à pH 8 et à pH 10 par rapport au type sauvage contrairement à H222R seule qui lui montre une stabilité accrue en milieu alcalin. Contrairement à la position interboucles ($i - 2$ et $i - 4$) où un pont salin est conservé chez différentes structures homologues avec une importante tolérance aux mutations et aux renversements de polarité, les deux autres positions impliquant des résidus histidine ne sont pas aussi bien conservées (Figure B.20). L'introduction du même genre de mutation H→R aux deux autres positions n'a pas eu le même effet thermostabilisant que H222R. Ces résultats démontrent les limites de

l'efficacité de la prédiction rationnelle sur une base comparaison par homologie ayant pour but de modifier les propriétés d'une enzyme dans une direction voulue.

6.3.7 Perspectives

La nouvelle enzyme lipolytique thermostable EstGtA2 a été rapportée et caractérisée ici pour la première fois. L'identification de motifs structuraux et l'utilisation de leur coévolution comme marqueur séquentiel de la nouvelle sous-famille N' (nouvelle famille XV) ont ouvert la voie pour l'élucidation du rôle que jouent ces éléments dans la thermostabilité des enzymes lipolytiques de cette nouvelle famille. L'étude combinatoire par « alanine-scanning » (R/K→A) a démontré que les résidus basiques impliqués dans la formation des ponts salins caractéristiques de la sous-famille N' jouent un rôle essentiel dans la stabilité conformationnelle et l'activité à haute température de EstGtA2. Une étude complémentaire intéressante serait de réaliser la même étude combinatoire, mais en substituant les partenaires acides (chargés négativement) pour leurs équivalents isostériques (D→N et E→Q) afin d'évaluer la contribution de l'empaquetage des chaînes latérales basiques et du pontage hydrogène et d'isoler la contribution électrostatique de ces mêmes ponts salins. En plus du remplacement isostère, le remplacement des résidus basiques pour des résidus glutamine (Q) permettrait de compléter l'étude combinatoire sur le rôle des ponts salins multiples dans la thermostabilité de EstGtA2 et d'évaluer quelles composantes de ces ponts salins déterminent la thermostabilité de cette nouvelle enzyme.

Il sera fondamental de résoudre la structure cristalline de EstGtA2 dans un avenir rapproché afin d'étudier davantage les éléments structuraux de sa stabilité, mais aussi pour comprendre et expliquer une observation expérimentale intéressante. En effet, la nouvelle EstGtA2 catalyse l'hydrolyse de la tributyrine alors que l'homologue MGL H-257 (PDB no. 3RM3), présentant 89 % d'identité de séquence semble en être incapable. L'alignement de la structure cristalline de EstGtA2 et celle de MGL H-257 pourrait permettre d'identifier une crevasse ou une pochette supplémentaire dans la structure de EstGtA2, ou peut être la présence d'un large résidu dans la pochette de liaison de MGL

H-257 empêchant la liaison de ce triglycéride à courte chaîne, laquelle ne peut être résolue dans le modèle d'homologie. Cette hypothèse reste à vérifier et pourrait expliquer la différence de spécificité entre ces deux enzymes homologues. La nouvelle EstGtA2 constitue un modèle intéressant pour l'étude du repliement des protéines globulaires pour différentes raisons : elle est formée d'une chaîne polypeptidique de longueur moyenne (249 résidus), est exprimée de façon soluble chez *E. coli*, elle montre une réversibilité de dénaturation thermique dépendante du pH et son activité catalytique est facilement détectable par spectrophotométrie avec différents substrats chromogéniques. La prochaine étape pour l'utilisation de EstGtA2 comme protéine modèle pour l'étude de la stabilité et du repliement est la résolution de sa structure cristalline.

CHAPITRE VII

CONCLUSION

Une nouvelle enzyme lipolytique thermostable, EstGtA2 de *G. thermodenitrificans* a été produite et caractérisée pour la première fois. Nous proposons d'élargir la classification originale en introduisant une nouvelle famille d'enzymes lipolytiques bactériennes, la famille XV à laquelle appartient EstGtA2 et d'utiliser la coévolution de ponts salins conservés comme marqueurs séquentiels pour l'assignation d'enzymes à différents sous-groupes. Nous avons montré que les résidus basiques impliqués dans la série de ponts salins conservés, marqueurs de la nouvelle sous-famille N' (famille XV), jouent un rôle essentiel dans la thermostabilité de EstGtA2 et qu'une certaine combinaison de ces mêmes résidus semble être impliquée dans son repliement. Il n'est cependant pas clair si ces résidus de ponts salins ont été sélectionnés pour leur rôle stabilisant dans la structure native de EstGtA2 ou s'ils peuvent jouer un rôle direct dans le repliement soit par des effets de répulsions électrostatiques et/ou par la formation de ponts salins transitoires spécifiques. Outre les résidus constituant le cœur hydrophobe d'une protéine, les autres éléments spécifiques d'un repliement seraient-ils discrètement localisés au niveau de la séquence? Différents patrons de ponts salins conservés au sein d'homologues appartenant à différents sous-groupes de la famille XV et XIII sont observés, ce qui suggère que si certains de ces résidus jouent un rôle spécifique dans le repliement, ils peuvent varier et se compenser au sein des enzymes d'une même famille. Nous avons aussi montré qu'un pont salin particulier liant deux boucles du site actif, conservé au sein de la famille XV et XIII, permet de contrôler la stabilité de EstGtA2 en fonction du pH. Ce pont a été utilisé comme cible pour améliorer la thermostabilité de EstGtA2 en milieu alcalin. La démonstration que des résidus chargés impliqués dans la formation de ponts salins peuvent être impliqués dans le repliement des protéines et que ceux-ci peuvent servir de cibles pour la modulation de la réponse au pH, ouvre la voie pour l'ingénierie d'autres enzymes lipolytiques pour différentes applications dans le futur.

RÉFÉRENCES BIBLIOGRAPHIQUES

1. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains. *Science* 181: 223-230.
2. Anson, M. L., and Mirsky, A. E. (1934) The Equilibria between Native and Denatured Hemoglobin in Salicylate Solutions and the Theoretical Consequences of the Equilibrium between Native and Denatured Protein. *Journal General Physiology* 17, (3): 393-408.
3. Arpigny J. L., and Jaeger K. E. (1999) Bacterial lipolytic enzymes: classification and properties. *Biochem J* 343: 177-183.
4. Backman, J., Schäfer, G., Wynds, L., and Bönisch, H. (1998) Thermodynamics and kinetics of unfolding of the thermostable trimeric adenylate kinase from the archaeon *Sulfolobus acidocaldarius*. *J. Mol. Biol* 284: 817-833.
5. Bae, E., and Phillips, G. N. (2004) Structures and analysis of highly homologous psychrophilic, mesophilic, and thermophilic adenylate kinases. *J Biol Chem* 279: 28202-28208. doi: 10.1074/jbc.M401865200.
6. Behe, M., Lattman, E.E., and Rose, G.D. (1991) The protein-folding problem: The native fold determines packing, but does packing determine the native fold? *Proc. Natl. Acad. Sci. USA*. 88(10): 4195-4199.
7. Ben-Naim, A. (1990) On the Role of Hydrogen-Bonds in Protein Folding and Protein Association *J. Phys. Chem* 95(3): 1444-1473.
8. Ben-Naim, A. (1989) Solvent-Induced Interactions: Hydrophobic and Hydrophilic Phenomena *Journal of Chemical Physics* 90(12): 7412-7525.
9. Ben-Naim, A. (1990) Solvent Induced Forces in Protein Folding, *Journal of Chemical Physics*, 94(17): 6893-6895.
10. Bernal, J. D. (1926) On the interpretation of X-ray, single crystal, rotation photographs. *Proc. R. Soc. Lond.* 113(763): 117-160.
11. Bernal, J. D., and Crowfoot, D. (1933) Crystalline phases of some substances studied as liquid crystals. *Trans. Faraday Soc.* 29: 1032-1049.

12. Bernal, J. D. (1953) The use of Fourier transforms in protein crystal analysis. *Proc. R. Soc. Lond.* 141(902): 71-85.
13. Bloom, J. D., Labthavikul, Sy. T., Otey, C. R., and Arnold, F. H. (2005) Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. USA* 103: 5869-5874.
14. Bolon, D. N., Voigt, C. A., and Mayo, S. L (2002) De novo design of biocatalysts. *Curr Opin. Struct. Biol* 6: 125-129.
15. Bornscheuer, U. T., and Pohl, M. (2001) Improved biocatalysts by directed evolution and rational protein design. *Curr Opin. Struct. Biol* 5:137-143.
16. Bornscheuer, U. T. (2002) Microbial carboxyl esterases: classification, properties and application in biocatalysis. *FEMS Microbiol Rev* 26: 73-81.
17. Bowie, J. U., and Sauer, R.T. (1989) Identifying determinant of folding and activity for a protein of unknown structure. *Proc. Natl. Acad. Sci. USA* 86, 2152.
18. Bowie, J. U., Reidhhar-Olson, J. F., Lim, W. A., and Sauer, R. T (1990) Deciphering the message in protein sequences: tolerance to amino acid substitution. *Science* 247: 1306-1310.
19. Bowie, J. U., Luthy, R., and Eisenberg, D. (1991) A method to identify protein sequences that fold into a known three-dimensional structure *Science* 253: 164-170.
20. Bowie, J. U. (2005) Solving the membrane protein folding problem *Nature* 438: 581-589.
21. Bragg, W. H. (1913) The reflections of X-rays by crystals. (II.). *Proc. R. Soc. Lond.* 89: 246-248.
22. Bragg, W. L. (1913) The structures of some crystals as indicated by their diffraction of X-rays. *Proc. R. Soc. Lond.* 89: 248-277.
23. Bragg, W. L., and Perutz, M. F. (1954) The structure of haemoglobin V1: Fourier projection on the 010 planes. *Proc. R. Soc. Lond* 225: 315-319.
24. Brenner, S., Barnett, L., Crick, F., and Orgel, A. (1961) The theory of mutagenesis. *J. Mol. Biol* 3: 121-124.
25. Brown, B. M., and Sauer, R. T. (1999) Tolerance of Arc repressor to multiple-alanine substitutions. *Proc. Natl. Acad. Sci. USA* 96: 1983-1988.

26. Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G. (2004) Funnels, pathways and the energy landscape of protein folding: A synthesis. *Proteins* 21: 167-195.
27. Chica, R. A., Doucet, N., and Pelletier, J. N. (2005) Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr. Opin. Struct. Biol* 16: 378-384
28. Chica, R. A., Moore, M. M., Allen, B. D., and Mayo, S. L. (2010) Generation of longer emission wavelength red fluorescent proteins using computationally designed libraries. *Proc. Natl. Acad. Sci. USA* 107: 20257-20262.
29. Chick, H., and Martin, C. J. (1910) On the heat-coagulation of proteins *J. Physiol* 40: 404.
30. Chick, H., and Martin, C. J. (1911) On the heat-coagulation of proteins: The action of hot water upon egg-albumen and the influence of acid upon reaction velocity. *J. Physiol* 43: 1-27.
31. Cooper, S., Khatib, F., Treuille, A., Barbero, J., Beenen, M., Leaver-Fay, A., Baker, D., Popović, Z., and Playesr, F. (2010) Predicting protein structure with a multiplayer online game. *Nature* 5: 756-760.
32. Creighton, T E. (1985) The problem of how and why proteins adopt folded conformation. *J. Phys. Chem.* 89: 2452-2459.
33. Crick, F. H. C. (1956) The theory of the method of isomorphous replacement for protein crystal. I. *Acta cryst* 9: 901-908.
34. Crick, F. (1958) On protein synthesis. The symposia of the society of experimental biology 12: 138-163.
35. Crick, F., Barnett, L., Brenner, S., and Watts-Tobin, R. J. (1961) General nature of the genetic code for proteins. *Nature* 192: 1227-1232.
36. Crick, F. (1988) What mad pursuit: A personal view of scientific discovery. Basic books, New York, USA.
37. Dahiyat, B. L., and Mayo, S. L (1996) Protein design automation. *Protein Sci* 5: 895-903.
38. Dahiyat, B. L., and Mayo, S. L (1997) De novo protein design: fully automated sequences selection. *Science* 278: 82-87.

39. Dantas, G., Kuhlman, B., Callender, D., Wong, M., and Baker, D. (2003) A large scale test for computational protein design: folding and stability of nine completely redesigned globular proteins. *J. Mol. Biol* 332: 449-460.
40. del Rio, J. C., Prinsen, P., and Guitiérrez, A. *Agricultural and food chemistry* 61: 1904-1913.
41. Del Vecchio, P., Graziano, G., Granata, V., Barone, G., Mandrich, L., Rossi, M., and Manco, G. (2002) Denaturing action of urea and guanidine hydrochloride toward two thermophilic esterases. *Biochem J* 367: 857-863.
42. Del Vecchio, P., Graziano, G., Granata, V., Barone, G., Mandrich, L., Rossi, M., and Manco, G. (2002) Temperature- and denaturant-induced unfolding of two thermophilic esterases. *Biochemistry* 41: 1364-1371.
43. Del Vecchio *et al.* (2009) Structural determinant of the high thermal stability *SsoPox* from the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Extremophiles* 13: 461-470.
44. Deutschman, W., and Dahlquist, F. W. (2001) Thermodynamic basis of the increased thermostability of CheY from the hyperthermophile *Thermotoga maritima*. *Biochemistry* 40: 13107-13113.
45. Dill, K. A. (1985) Theory of the folding and stability of globular proteins. *Biochemistry* 24: 1501-1509.
46. Dill, K. A. (1990) Dominant Forces in Protein Folding, *Biochemistry* 29(31): 7133-7155.
47. Dill, K. A. (1997) From Levinthal to pathways to funnels. *Nat. Struct. Biol* 4: 10-19.
48. Dill, K. A., Ozkan, S. B., Wei, T. R., Chodera, J. D., and Voelz, V. A. (2007) The protein folding problem: when it will be solved? *Curr. Opin. Struct. Biol* 17: 342-347.
49. Dill, K. A., Ozkan, S. B., Shell, M. S., and Weikl, T. R. (2008) The protein folding problem. *Annual Review of Biophysics* 37: 289-316.
50. Diaz P, Prim N, and Javier Pastori FI (1999) Direct fluorescence-based lipase activity assay. *Biotechniques* 27: 696-700.

51. Edsall, J. T. (1995) Hsien Wu and the first theory of protein denaturation. *Advances in Protein Chemistry* 46: 1-5.
52. Eisenberg, D., and McLachlan A. D. (1986) Solvation energy in protein folding and binding, *Nature* 319: 199-203.
53. Eisenberg, D., Weiss, R. M., Terwilliger, T.C., and Wilcox, W. (1982) Hydrophobic moments and protein structure. *Faraday. Symp. Chem. Soc.* 17: 109-120.
54. Elcock, A. H. (1998) The stability of salt bridges at high temperatures: implication for hyperthermophilic proteins. *J. Mol. Biol* 284: 489-502.
55. Fersht, A. (1999) *Structure and Mechanism in Protein Science* W. H. Freeman and Company, New York.
56. Fischer, D., and Eisenberg, D. (1996) Protein fold recognition using sequence-derived prediction. *Protein Science* 5: 947-955.
57. Fischer, M., and Pleiss, J. (2003) The lipase engineering database: a navigation and analysis tool for protein families. *Nucleic Acids Res* 31: 319-321.
58. Fogolari, F., Brigo, A., and Molinari, H. (2002) The Poisson-Boltzman equation for biomolecular electrostatics: a tool for structural biology. *Journal of molecular recognition* 15: 377-392.
59. Fojan, P., Jonson, P. H., Petersen, M. T., and Petersen, S. B. (2000) What distinguishes an esterase from a lipase: a novel structural approach. *Biochimie* 82: 1033-1041.
60. Gao, B., Su, E., Lin, J., Jiang, Z., Ma, Y., *et al.* (2009) Development of recombinant *Escherichia coli* whole-cell biocatalyst expressing a novel alkaline lipase-coding gene from *Proteus* sp. for biodiesel production. *J Biotechnol* 139: 169-75.
61. Gilson, M. K. (1995) Theory of electrostatic interactions in macromolecules. *Curr Opin Struct Biol* 5: 216-223.
62. Gordon, D. B., Marshall, S. A., and Mayo, S. L (1999) Energy function for protein design. *Curr. Opin. Struct. Biol* 9: 509-513.

63. Gupta R, Gupta N, and Rath P (2004) Bacterial lipases: an overview of production, purification and biochemical properties. *Appl Microbiol Biotechnol* 64: 763-781.
64. Haschemeyer, A., Persell, R. (1973) Kinetic studies of amino acid uptake and protein synthesis in liver of temperature acclimated toad fish. *Biological Bulletin* 145: 472-481.
65. Hilbert, M., and López, P. (2011) The world's technological capacity to store, communicate and compute information. *Science*. 1;332(6025): 60-65.
66. Hildebrand, A., Remmert, M., Biegert, A., and Söding, J. (2009) Fast and accurate automatic structure prediction with HHpred. *Proteins* 9:128-132.
67. Jacobsen, C. F., and Linderstørn-Lang, K. (1949) Salt linkages in proteins. *Nature* 164: 411-412.
68. Jaeger, K. E., Dijkstra, B. W., and Reetz, M. T. (1999) Bacterial biocatalysts: molecular biology, three-dimensional structures, and biotechnological applications of lipases. *Annu Rev Microbiol* 53: 315-351.
69. Jaeger, K. E., and Eggert, T. (2002) Lipases for biotechnology. *Curr Opin Biotechnol* 13:390-397.
70. Jaeger, K. E., and Reetz, M. T. (1998) Microbial lipases form versatile tools for biotechnology. *Trends Biotechnol* 16: 396-403.
71. Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Rothlisberger, D., Zanghellini, and A., Baker, D (2008) De novo computational design of retro-aldol enzymes. *Science* 7: 1387-91.
72. Kauzmann, W. (1959) Some Factors in the Interpretation of Protein Denaturation. *Advances in Protein Chemistry* 14: 1-63.
73. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C. (1958) A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature* 181: 662-666.
74. Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davis, D. R., Phillips, D. C., and Shore, V.C. (1960) A three-dimensional Fourier synthesis at 2 Å resolution. *Nature* 185: 422-427.

75. Khatib, F., Cooper, S., Tyka, M. D., Makedon, I., Popovic, Z., Baker, D., and Players, F. Algorithm discovery by protein folding game players. *Proc. Natl. Acad. Sci. USA* 108: 18949-18953.
76. Kirk, O., Borchert, T. V., and Fuglsang, C. C (2002) Industrial enzymes applications. *Current opinion in biotechnology* 13: 345-351.
77. Knöchel, T., Pappenberger, A., Jansonius, J. N., and Kirschner, K. (2002) The crystal structure of indoleglycerol-phosphate synthase from *Thermotoga maritima*. Kinetic stabilization by salt bridges. *J Biol Chem* 277: 8626-834.
78. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., and Baker, D. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science* 302: 1364-1368.
79. Kulik, H. J., Luehr, N., Ufimtsev, I. S., and Martinez, T. J. (2012) Ab initio quantum chemistry for protein structures. *J. Phys. Chem* 116: 12501-12509.
80. Lattman, E. E., and Rose, G.D. (1993) Protein folding-What's the question? *Proc. Natl. Acad. Sci. USA* 90(2): 439-441.
81. Levinson, M., van der Oost, J., and Kengen, S. W. M. (2009) Carboxylester hydrolases from hyperthermophiles. *Extremophiles* 13: 567-581.
82. Levinthal, C. (1968) Are there pathways for protein folding? *Extrait du journal de chimie physique* 65 :44.
83. Linderstrøm-Lang K. (1951) Proteins and enzymes. Lane medical lectures. Oxford University Press, Oxford, UK.
84. Lloyd, S. (2000) Ultimate physical limits to computation. *Nature*. 31;406(6799): 1047-1054.
85. Luke, K. A., Higgins, C. L., and Stafshide, P. W. (2007) Thermodynamic stability and folding of proteins from hyperthermophilic organisms. *FEBS Journal* 274: 4023-4033.
86. Marqusee, S. and Sauer, R. T (1994) Contribution of a hydrogen bond/salt bridge network to the stability of secondary and tertiary structure in λ repressor. *Protein Science* 3: 2217-2225.
87. Marshall, S. A., and Mayo, L. S. (2001) Achieving stability and conformational specificity in designed proteins via binary patterning. *J. Mol. Biol* 305: 619-631.

88. Marshall, S. A., Morgan, C. S., and Mayo, L. S. (2002) Electrostatics significantly affect the stability of designed homoeodomains variants. *J. Mol. Biol* 316: 189-199.
89. Mirsky, A. E., and Pauling, L. (1936) On the Structure of Native, Denatured and Coagulated Proteins *Proc. Natl. Acad. Sci. USA* 22(7): 439-447.
90. Mooers, B. H., Baase, W. A., Wray, J. W., and Matthews, B. W. (2009) Contribution of all 20 amino acids at site 96 to the stability and structure of T4 lysozyme. *Protein Science* 18: 871-880.
91. Nardini, M., and Dijkstra, B. W. (1999) Alpha/bêta hydrolase fold enzymes: the family keeps growing. *Curr Opin Struct Biol* 9: 732-737.
92. Nojima, H., Ikai, A., Oshima, T., and Noda, H. (1977) Reversible thermal unfolding of thermostable phosphoglycerate kinase. Thermostability associated with mean zero enthalpy change. *J. Mol. Biol* 116: 429-442.
93. Ollis D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., *et al.* (1992) The alpha/beta hydrolase fold. *Protein Eng* 5: 197-211.
94. Pace, N. C. (1996) Forces contributing to the conformational stability of proteins. *FASEB J* 10: 75-83.
95. Pace, N. C., Shirley, B. A., and Thompson J. A. Measuring the conformational stability of proteins. In *Protein Structure-A. Practical Approach*; Creighton, T. E., Ed.; IRL: Oxford, 1989; Chapter 13.
96. Pace, N. C. (1975) The stability of globular proteins. *Critical review in biochemistry and molecular biology*.
97. Pace, N. C., and Grimsley, R. G. (2011) Contribution of hydrophobic interactions to protein stability. *J. Mol. Biol* 408: 514-528.
98. Pauling, L. (1948) *The Nature of Chemical Bond*, 2nd Edition, Cornell University Press, Ithaca.
99. Pauling, L. (1960) *The Nature of Chemical Bond*, 3rd Edition, Cornell University Press, Ithaca.
100. Pauling, L., and Corey, R. B. (1951) Atomic Coordinates and Structure Factors for Two Helical Configurations of Polypeptide Chains *Proc. Natl. Acad. Sci. USA* 37(5): 235-248.

101. Pauling, L., and Corey, R. B. (1951) The Pleated Sheet, A New Layer Configuration of Polypeptide Chains *Proc. Natl. Acad. Sci. USA* 37(5): 251-256.
102. Pauling, L., and Corey, R. B. (1951) The Structure of Fibrous Proteins of the Collagen-Gelatin Group *Proc. Natl. Acad. Sci. USA* 37(5): 272-281.
103. Pauling, L., and Corey, R. B. (1951) Configurations of Peptide Chains with Favored Orientations around Single Bonds: Two New Pleated Sheets. *Proc. Natl. Acad. Sci. USA* 37(5): 729-740.
104. Perutz, M. F. (1978) Electrostatic effects in proteins. *Science* 29: 1187-1191.
105. Pezzullo, M., Del Vecchio, P., Mandrich, L., Nucci, R., Rossi, M., and Manco, G. (2013) Comprehensive analysis of surface charged residues involved in thermal stability in *Allicyclobacillus acidocaldarius* esterase 2. *Protein. Eng. Des. Sel* 26: 47-58.
106. Plotkin, S. S., and Onuchic S. (2002) Understanding protein folding with energy landscape theory. *Quarterly Reviews of biophysics* 35: 111-167.
107. Privett, H. K., Kiss, G., Lee, T. M., Bloomberg, R., Chica, R. A., Thomas, L. M., Hilvert, D., Houk, K. N., and Mayo, S. L. (2012) Iterative approach to computational enzyme design. *Proc. Natl. Acad. Sci. USA* 109: 3790-3795.
108. Ridgen, D. (2009) From protein structure to function with bioinformatics. Springer, Liverpool, UK.
109. Sauer, R. T., Milla, M. E., Waldburger, C. D., Brown, B. M., and Schildbach, J. F. (1996) Sequence determinants of folding and stability for the P22 Arc repressor dimer. *FASEB* 10: 42-48.
110. Rose, G. D., Flemming, P. J., Banavar, J. R., and Maritan, A. (2006) A backbone-based theory of protein folding. *Proc. Natl. Acad. Sci. USA* 143: 16623-16633.
111. Schellmann, J. A. (1955) The Thermodynamics of Urea Solutions and the Heat of Formation of the Peptide Hydrogen Bond. *Comptes Rendus des Travaux du Laboratoire Carlsberg, Serie Chimique*. 29(14-15): 223-230.
112. Schellmann, J. A. (1955) The Stability of Hydrogen-Bonded Peptide Structures in Aqueous Solution. *Comptes Rendus des Travaux du Laboratoire Carlsberg, Serie Chimique* 29(14-15): 230-259.

113. Schaeffer, R. D., and Dagget, V. (2011) Protein fold and protein folding. *Prot. Eng. Des. Sel* 24 : 11-19.
114. Sharma D, Sharma B, and Shukla AK (2011) Biotechnological approach tomicrobial lipase: a review. *Biotechnology* 10: 23-40.
115. Sharma R, Chisti Y, and Banerjee UC (2001) Production, purification, characterization, and applications of lipases. *Biotechnol Adv* 19: 627-662.
116. Simons, K. T., Bonneau, R., Ruczinski, I., and Baker, D. (1999) Ab initio protein structure prediction of CASP III targets using Rosetta. *Proteins* 3: 171-176.
117. Stetter, K. O. (2006) History of discovery of the first hyperthermophiles. *Extremophiles* 10: 357-362.
118. Tanford, C. (1968) Protein denaturation: Part A. Characterization of de denaturated state; Part B. The transition from native to denturated state. *Adv. Protein. Chem* 23:121.
119. Tanford, C. (1970) Protein denaturation: Part C. Theoretical model for the mechanism of denaturation. *Adv. Protein. Chem* 24:1.
120. Tanford, C., and Reynold, J. (2003) Nature's Robots, A History of Proteins Oxford University Press, Oxford.
121. Tissot, A. C., Vuilleumier, S., and Fersht, A. R. (1996) Importance of two buried salt bridges in the stability and folding pathways of Barnase. *Biochemistry* 35: 6786-6794.
122. Verger R (1997) Interfacial activation of lipases: facts and artifacts. *Tibtech* 15: 32-38.
123. Vetriani C, Maeder DL, Tolliday N, Yip KS, Stillman TJ, *et al.* (1998) Protein thermostability above 100 degrees C: a key role for ionic interactions. *Proc Natl Acad Sci USA* 95: 12300-12305.
124. Vijayakumar, M., and Zhou, H-X. (2001) Salt bridges stabilized the folded structure of Barnase. *J. Phys. Chem* 105: 7334-7340.
125. Vizcarra, C. L., and Mayo, S. L (2005) Electrostatic in computational protein design. *Curr. Opin. Struct. Biol* 9: 622-626.

126. Waldburger, C. D., Schildbach, J. F., and Sauer, R. T (1995) Are buried salt bridges important for protein stability and conformational specificity? *Nat. Struct. Biol* 2: 122-128.
127. Warshel, A. (1981) Electrostatic basis of structure-function correlation in proteins. *Acc. Chem. Res* 14: 284-290.
128. Warshel, A., and Papazyan, A. (1998) Electrostatic effects in macromolecules: fundamental concepts and practical modeling. *Curr. Opin. Struct. Biol* 8: 211-217.
129. Wassenberg, D., Welker, C., and Jaenick, R. (1999) Thermodynamics of the unfolding of the cold-shock protein from *Thermotoga maritima*. *J. Mol. Biol* 289: 187-193.
130. Watson, J. D., and Crick, F. H. C. (1953) Molecular structure of nucleic acid: A structure of deoxyribose nucleic acid. *Nature*. 171: 737-738.
131. Watson, J. D., and Crick, F. H. C. (1953) Genetical implications of the structure of deoxyribonucleic acid. *Nature*. 171: 964-967.
132. Zwanzig, R., Szabo, A., and Bagchi, B. (1992) Levinthal's paradox *Proc. Natl. Acad. Sci. USA* 89: 20-22.

ANNEXE A

DIESELZYMES: DEVELOPMENT OF A STABLE AND METHANOL TOLERANT LIPASE FOR BIODIESEL PRODUCTION BY DIRECTED EVOLUTION

Le présent chapitre est présenté sous forme d'un texte résumé en français. Ce chapitre traite des résultats issus d'une collaboration de recherche effectuée lors d'un stage réalisé à l'Université de Californie Los Angeles durant l'année 2010. Les résultats ont fait l'objet d'une publication dans le journal scientifique *Biotechnology for Biofuels*. L'article intégral publié peut être consulté en ligne suivant le lien ci-dessous.

Tyler P. Korman, Bobby Sahachartsiri, **David M. Charbonneau**, Grace Huang, Marc Beauregard and James U. Bowie. 2013. Dieselzymes: Development of a stable and methanol tolerant lipase for biodiesel production by directed evolution. *Biotechnology for Biofuels* **6**:70 doi:10.1186/1754-6834-6-70.

<http://www.biotechnologyforbiofuels.com/content/pdf/1754-6834-6-70.pdf>

A.1 Contribution des auteurs

Tyler P. Korman a conçu les expériences et réalisé les expériences biochimiques, a élucidé les structures cristallines et écrit le manuscrit. David M. Charbonneau et Tyler P. Korman ont développé et mis au point la nouvelle méthode de criblage. David M. Charbonneau, Bobby Sahachartsiri et Grace Huang ont participé à la génération et au criblage des mutants. Marc Beauregard a participé à l'analyse des résultats. James U. Bowie a mis en place et financé l'étude, il a participé à sa conception, à sa coordination et à l'écriture du manuscrit.

A.2 Résumé de l'article

Les biodiesels sont des carburants alternatifs composés d'esters méthyliques d'acides gras habituellement produits par des réactions de transestérification de triglycérides et de méthanol en conditions alcalines. Certaines lipases sont d'intéressants biocatalyseurs pour la synthèse de biodiesel et montrent plusieurs avantages en comparaison avec les méthodes traditionnelles par transestérification acide ou basique. Par contre, les lipases naturelles sont souvent rapidement inactivées par les hautes concentrations de méthanol utilisées pour la synthèse de biodiesel. La lipase de *Proteus mirabilis* (PML) est un candidat intéressant pour la production de biodiesel puisqu'elle produit de hauts rendements d'esters méthyliques d'acides gras, mais montre une faible stabilité et résistance au méthanol. Nous avons utilisé l'évolution accélérée pour améliorer la résistance au méthanol et la thermostabilité de PML, ce qui a mené à un variant contenant 13 mutations, appelé Dieselzyme 4, qui montre une thermostabilité et une résistance au méthanol supérieures au type sauvage de PML, avec un temps d'inactivation à 50 °C de plus 30 fois celle du type sauvage et de plus de 50 fois en présence de 50 % méthanol. Le mutant Dieselzyme 4 immobilisé conserve son habilité pour la synthèse de biodiesel après plusieurs cycles de réutilisation et montre une longévité grandement supérieure à celle du type sauvage de PML ou encore de la lipase vedette utilisée en industrie (BCL) de *Burkholderia cepaica*. Les mutations sélectionnées au cours des rondes d'évolution accélérée (N17S, R33T, L64I, A70T, G181C, G202E, K208N, F225L, S238C, I255F, G266S, D270N et Q277L) montrent un accroissement général de l'hydrophilicité de PML. La comparaison de la structure cristalline de PML (PDB : 4GW3) et du prototype Dieselzyme 4 (PDB : 4HS9) a révélé que les nouvelles interactions stabilisantes dans Dieselzyme 4 sont principalement des nouveaux ponts hydrogènes et des ponts salins, ce qui suggère que les interactions polaires peuvent être particulièrement stabilisantes dans l'environnement diélectrique réduit utilisé pour la synthèse de biodiesel. Le prototype Dieselzyme 4 constitue la lipase la plus robuste à ce jour pour la synthèse de biodiesel et est en attente de brevet.

A.3 Cinquième article scientifique

Dieselzymes: Development of a stable and methanol tolerant lipase for biodiesel production by directed evolution

Tyler P. Korman¹, Bobby Sahachartsiri¹ David M. Charbonneau², Grace Huang¹, Marc Beauregard² and James U. Bowie^{1*}

1. Department of Chemistry and Biochemistry, UCLA-DOE Institute of Genomics and Proteomics, Molecular Biology Institute, University of California Los Angeles, Los Angeles, USA
2. Département de chimie-biologie, Université du Québec à Trois-Rivières, Québec, Canada

James U. Bowie

*Corresponding author :

Email : bowie@mbi.ucla.edu

Keywords: Lipase, Biodiesel, *Proteus mirabilis*, *Proteus* sp. K107, Directed evolution, Alcohol tolerance

Published in: *Biotechnology for Biofuels* 2013, **6**:70 doi:10.1186/1754-6834-6-70

Original article available online at:

<http://www.biotechnologyforbiofuels.com/content/pdf/1754-6834-6-70.pdf>



RESEARCH

Open Access

Dieselzymes: development of a stable and methanol tolerant lipase for biodiesel production by directed evolution

Tyler P Korman¹, Bobby Sahachartsiri¹, David M Charbonneau², Grace L Huang¹, Marc Beauregard² and James U Bowie^{1*}

Abstract

Background: Biodiesels are methyl esters of fatty acids that are usually produced by base catalyzed transesterification of triacylglycerol with methanol. Some lipase enzymes are effective catalysts for biodiesel synthesis and have many potential advantages over traditional base or acid catalyzed transesterification. Natural lipases are often rapidly inactivated by the high methanol concentrations used for biodiesel synthesis, however, limiting their practical use. The lipase from *Proteus mirabilis* is a particularly promising catalyst for biodiesel synthesis as it produces high yields of methyl esters even in the presence of large amounts of water and expresses very well in *Escherichia coli*. However, since the *Proteus mirabilis* lipase is only moderately stable and methanol tolerant, these properties need to be improved before the enzyme can be used industrially.

Results: We employed directed evolution, resulting in a *Proteus mirabilis* lipase variant with 13 mutations, which we call Dieselzyme 4. Dieselzyme 4 has greatly improved thermal stability, with a 30-fold increase in the half-inactivation time at 50°C relative to the wild-type enzyme. The evolved enzyme also has dramatically increased methanol tolerance, showing a 50-fold longer half-inactivation time in 50% aqueous methanol. The immobilized Dieselzyme 4 enzyme retains the ability to synthesize biodiesel and has improved longevity over wild-type or the industrially used *Burkholderia cepacia* lipase during many cycles of biodiesel synthesis. A crystal structure of Dieselzyme 4 reveals additional hydrogen bonds and salt bridges in Dieselzyme 4 compared to the wild-type enzyme, suggesting that polar interactions may become particularly stabilizing in the reduced dielectric environment of the oil and methanol mixture used for biodiesel synthesis.

Conclusions: Directed evolution was used to produce a stable lipase, Dieselzyme 4, which could be immobilized and re-used for biodiesel synthesis. Dieselzyme 4 outperforms the industrially used lipase from *Burkholderia cepacia* and provides a platform for still further evolution of desirable biodiesel production properties.

Keywords: Lipase, Biodiesel, *Proteus mirabilis*, *Proteus* sp. K107, Directed evolution, Alcohol tolerance

Background

Biodiesel is a well-validated transportation fuel that is an attractive alternative to petrodiesel [1]. Biodiesel burns cleaner, releases less CO₂ to the atmosphere, is biodegradable and can be obtained from renewable sources [1,2]. Moreover, biodiesel can be used in existing diesel engines and is compatible with current fuel distribution

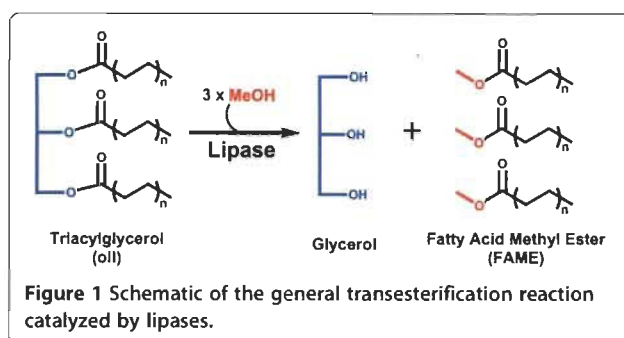
infrastructure. Finding new sources of biodiesel and improved production methods is therefore an important goal in achieving diversification of energy resources [2].

Biodiesel is typically composed of a mixture of fatty acid methyl esters (FAME) (Figure 1), although other esters can be used [3]. Biodiesel is usually synthesized via the base-catalyzed transesterification of triacylglycerol (TAG) oils with methanol (MeOH) or ethanol (EtOH). In the base-catalyzed reaction, the presence of water leads to the production of free fatty acids (FFAs), which are a dead-end product. Moreover, any FFAs present in

* Correspondence: bowie@mbi.ucla.edu

¹Department of Chemistry and Biochemistry, UCLA-DOE Institute of Genomics and Proteomics, Molecular Biology Institute, University of California, Los Angeles, USA

Full list of author information is available at the end of the article



the oil cannot be converted to methyl esters. FFA contamination requires additional purification and also leads to emulsions that complicate reaction clean up. In addition, glycerol, the other value-added reaction product, becomes contaminated with salt and base, increasing purification costs and lowering its value [1].

Lipases are a family of enzymes that are interesting as an alternative catalyst for biodiesel production [4]. Like chemical catalysts, lipases can catalyze the transesterification of TAG with short chain alcohols at significant rates. However, unlike the base-catalyzed reaction, some lipases can also convert FFAs to FAMES even in the presence of high levels of water [4,5]. Moreover, the lipase catalysts can be easily separated from the glycerol product, thereby eliminating glycerol contamination and increasing its potential as a value-added product to lower the net cost of biodiesel production [6,7].

An obvious drawback to the use of lipases is the high cost of the catalyst relative to a simple base such as NaOH [7,8]. The initial investment cost of a lipase catalyst could be greatly mitigated, however, if it could be used for long periods of time to generate large quantities of biodiesel. Unfortunately, the presence of high concentrations of MeOH used for biodiesel synthesis can cause rapid loss of lipase activity due to alcohol-induced inactivation [9].

Various approaches have been employed to increase lipase lifetimes for biodiesel synthesis [10]. Process engineering approaches have included enzyme immobilization, the use of co-solvents, or the step-wise feeding of alcohol so the concentration remains low [11]. However, the use of co-solvents further increase costs while step-wise feeding of MeOH is difficult on an industrial scale. In addition, there have been efforts to identify natural lipases that are inherently more tolerant of the harsh conditions employed for biodiesel production [12]. To our knowledge, rational design or directed evolution approaches have not been used to improve natural lipases specifically for this non-natural role.

Directed evolution and rational design of existing lipases have the potential to produce long lasting re-engineered enzymes with specific properties [13]. For

example, previous reports have shown that many α/β hydrolase fold enzymes, including lipases, have been successfully reengineered by directed evolution for improved thermostability as well as tolerance to a variety of polar solvents such as dimethylsulfoxide (DMSO), dimethylformamide (DMF), and acetonitrile (MeCN) [14-18]. To date, there are no reports of the reengineering of a lipase specifically for improved tolerance to short-chain alcohols such as MeOH and EtOH. The development of a lipase highly resistant to MeOH inactivation would be especially useful as a catalyst for the economical synthesis of biodiesel.

Many microbial lipases have been identified as potential enzyme catalysts for biodiesel synthesis with a wide range of stabilities and catalytic efficiency [12]. In spite of the large number of identified microbial lipases, most are poor targets for directed evolution methods because they require chaperones and/or post-translational modifications specific to the host organism [19]. As a result most microbial lipases can only be produced in their native host, hampering engineering efforts. As an example, one of the most widely used industrial lipases, LipA from *Burkholderia cepacia*, is highly active and tolerant to short-chain alcohols but requires a chaperone for proper folding and does not express solubly in *Escherichia coli* [19].

Recently, an effective biodiesel-producing lipase from *Proteus sp.* K107 was identified that can be expressed solubly at high levels in *E. coli* [20]. The *Proteus sp.* K107 lipase is 100% identical to a lipase from *Proteus mirabilis* so we will refer to it as *Proteus mirabilis* lipase (PML). PML belongs to the *Proteus*/psychrophilic subfamily of I.1 lipases which lack a leader sequence and a disulfide bond present in other family I.1 and I.2 lipases [21]. PML was shown to be tolerant to short-chain alcohols such as MeOH and EtOH and to catalyze the synthesis of fatty acid methyl esters (biodiesel) in the presence of high concentration of water. Highlighting the usefulness of PML as a potential low-cost catalyst for biodiesel production [22], a recombinant *E. coli* strain overexpressing PML could also be used as a whole-cell biocatalyst for FAME synthesis [20]. Unfortunately, PML is irreversibly inactivated by incubation at over 50% MeOH, a common weakness of lipases [23]. In this paper we report the development of a PML variant with significant increases in thermostability and tolerance to MeOH that retains nearly wild-type activity at ambient temperature. A high-resolution crystal structure of the optimized lipase provides insight into adaptations in PML leading to heat and MeOH tolerance.

Results and discussion

Dieselzyme 1: generation of disulfide mutant increases stability

The introduction of disulfide bonds is a common strategy to improve enzyme stability that has successfully been

applied to other lipases [24,25]. Based on sequence analysis, we hypothesized that we could stabilize the PML by the introduction of a disulfide bond. Homologous lipases from *Pseudomonas aeruginosa* and *Burkholderia cepacia* (42% and 38% ID to PML respectively) contain a single disulfide bond between residues 181 and 238 (PML numbering) that is not conserved in PML [26,27]. The PML wild-type structure showing the location of the corresponding disulfide bond in *B. cepacia* (1OIL) can be seen in Figure 2A. Due to the structural conservation of this region and the proximity of G181 and S238 in PML, a disulfide bond was introduced between residues G181 and S238 by mutation to cysteines (Figure 2A). We refer to this mutant as Dieselzyme 1.

The effect of the introduced disulfide bond on resistance to thermal inactivation was determined by monitoring the residual lipase activity after heat treatment. Samples were incubated for 1 hour between 37 and 60°C and the residual activity was measured at 25°C. As shown in Figure 2B, the introduction of a single disulfide in Dieselzyme 1 significantly stabilized PML, increasing

the half-inactivation temperature ($IT_{1/2}$) from 37°C for the wild-type enzyme to 48°C for the Dieselzyme 1. Also, when incubating at a constant 50°C, the half-life increased from less than 15 min for wild-type to ~75 min for Dieselzyme 1 (Additional file 1: Figure S1). Interestingly, Dieselzyme 1 displays the same temperature inactivation profile as the native *Burkholderia cepacia* lipase (BCL) which contains a single disulfide in the same position (Figure 2B, Additional file 1: Figure S1). Analysis of the crystal structure (below) shows clear electron density for the disulfide bond (Additional file 2: Figure S2A) suggesting the improvement in thermostability is a direct result of disulfide bond formation.

While thermostability is an important property that contributes to the industrial usefulness of an enzyme, a more thermostable enzyme does not necessarily mean that an enzyme will be more tolerant to organic solvents, especially relatively polar, water miscible solvents such as methanol and ethanol that are used in biodiesel synthesis. Indeed, work with *Pseudomonas fragi* lipase has shown that increasing thermostability does not necessarily lead to

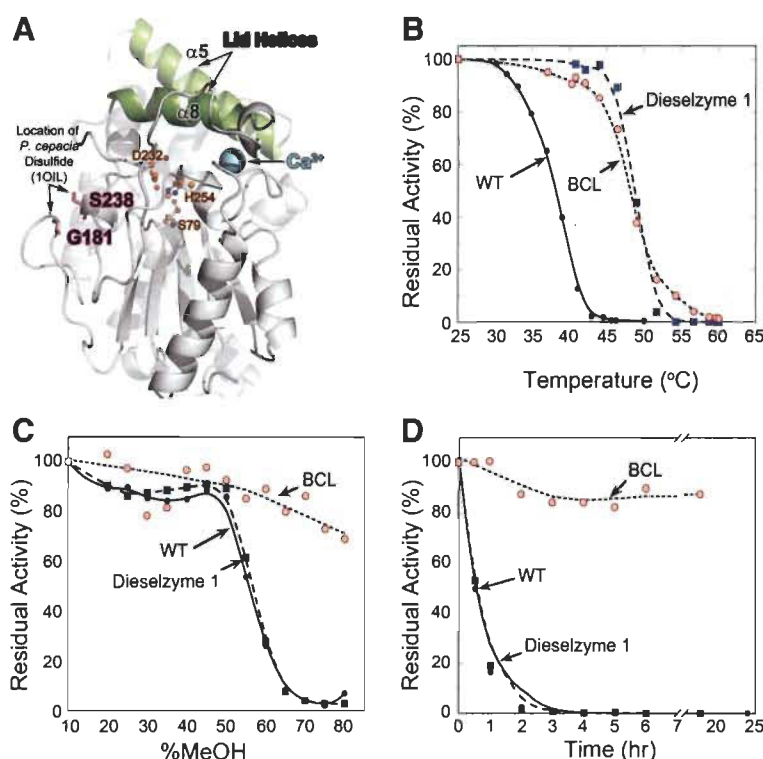


Figure 2 Effect of introduced disulfide bond on thermal and methanol induced inactivation. (A) Structure of *P. mirabilis* lipase. Lid helices are shown in green, bound Ca^{2+} in blue, and catalytic triad as orange ball and stick. The location of the equivalent disulfide bond found in the *B. cepacia* lipase but absent in PML is labeled and shown as sticks. (B) Thermal inactivation as a function of temperature. Residual activity was measured after incubation for 1 hour. (C) Inactivation as a function of methanol concentration. Residual activity was measured after incubation for 2 hours after dilution to 10% methanol. (D) Time course of inactivation by 50% methanol. Samples were diluted to 10% methanol prior to assaying for residual activity. Results are the average of three independent experiments. Error-bars are omitted but are less than 5% in all cases. Lines are shown for clarity and represent the best-fit to the data points. WT PML (black circles), BCL (red open circles), and Dieselzyme 1 (blue squares).

improved solvent tolerance [28]. Additionally, many thermostable lipases such as the *Thermomyces lanuginosus* lipase display only modest alcohol tolerance [29]. The effect of methanol on wild-type PML and Dieselzyme 1 was monitored by measuring the residual activity after incubation with various concentrations of methanol for 2 hours at 25°C. The introduced disulfide bond in Dieselzyme 1 does not have an effect on the methanol tolerance of PML as the methanol inactivation profiles for wild-type PML and Dieselzyme 1 are identical (Figure 2C and 2D). Thus, consistent with earlier work, it appears that the mechanism of methanol inactivation is different from heat inactivation. We therefore employed directed evolution to improve the methanol tolerance Dieselzyme 1.

Dieselzymes 2–4: directed evolution

To screen for mutants with improved methanol and heat tolerance, we developed a rapid colony assay for stability screening (Figure 3). Following error-prone PCR, mutant library colonies were lifted onto filter paper and protein expression was induced by placing the colonies face up on LB-agar containing 1 mM IPTG for 3 hours. The lipase variants expressed in each colony were then tested for their resistance to inactivation by incubating the filters in heat and methanol. After removal of the methanolic incubation solution, mutants that retained

significant residual activity could be identified by overlaying with a solution of 1-naphthyl palmitate and Fast Blue B dye suspended in agar, leading to the formation of a dark purple/brown colony when active lipase is present.

To screen for improved methanol tolerance, we started with the stable disulfide mutant Dieselzyme 1 as the parent for error-prone PCR. In the first round, the mutant library was incubated with 50% methanol at 45°C for 30 minutes prior to screening for residual activity. In the second round, the mutant library was incubated with 65% methanol at 45°C for 45 minutes. Finally, the third round of screening was performed in 65% methanol at 47°C for 90 minutes. During each round of screening, only the mutant clones displaying residual activity within the first ten minutes were picked and confirmed in a second colony screen as above. Typically, 10–15 mutants out of ~20,000 were selected for further validation during each round of screening.

The mutant lipases positive for increased methanol resistance after each round were purified and tested for their tolerance to methanol by measuring residual activity after a 16 hr incubation in 70% methanol at 25°C. Beneficial mutations identified in each round were combined iteratively via site-directed mutagenesis and then tested for increased methanol tolerance. The best recombinant from each round was then used as the parent

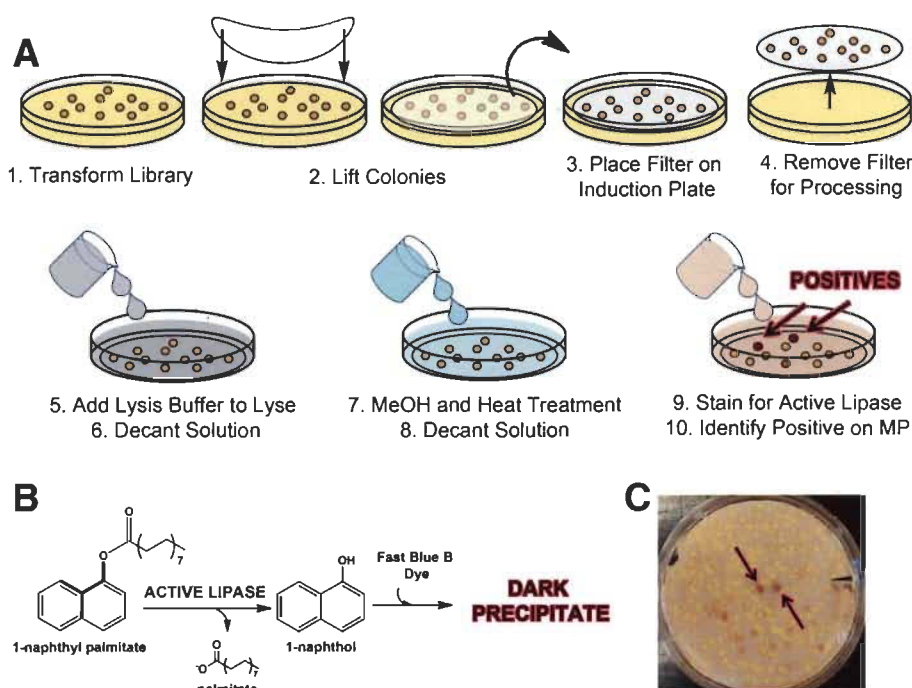


Figure 3 Schematic of the screening protocol for improved methanol tolerance by directed evolution of PML. (A) Steps for colony lift assay and staining protocol. (B) Hydrolysis of 1-naphthyl palmitate and detection of 1-naphthol by fast blue to assay colonies for residual activity after exposure to methanol and heat. (C) A representative filter following step 9 in (A). Black arrows point to potential positives identified on the filter.

for the next round of error-prone PCR. After introduction of the disulfide bond (Dieselzyme 1) and 3 rounds of random mutagenesis with site-directed recombination (Dieselzymes 2–4), a mutant (Dieselzyme 4) with 13 amino acid changes (Table 1) was identified. Dieselzyme 4 retained more than 80% of its residual activity after incubation with 70% methanol for 16 hours and displayed further improved thermostability compared to the disulfide mutant Dieselzyme 1 (see below).

Characterization of thermostability and alcohol tolerance of evolved PML mutants

To evaluate the full extent of stabilization after random mutagenesis, the residual activity of wild-type PML was compared with each Dieselzyme variant as a function of temperature or methanol concentration as shown in Figures 4 and 5 respectively. The thermal inactivation profile for the disulfide bond mutant Dieselzyme 1 and Dieselzyme 2 are identical. However, in Dieselzyme 3 and Dieselzyme 4, the added mutations have a beneficial effect on thermostability with an apparent $t_{1/2}$ after 1 hr of 55°C for both Dieselzyme 3 and Dieselzyme 4, which is an increase of 17°C compared to wild-type and ~5°C higher than Dieselzyme 1 (Figures 1B, 4A). In addition, when the mutant lipases were incubated at a constant 50°C, the final mutant Dieselzyme 4 also showed a further improvement over Dieselzyme 3 with an inactivation half-life of ~7 hours (Figure 4B) which is more than a 30-fold improvement over wild-type (less than 15 min). Dieselzyme 3 and Dieselzyme 4 are also more resistant to thermal inactivation than the industrial lipase from *B. cepacia* described above.

In addition to thermostability, the methanol tolerance of PML was also significantly improved over four rounds of screening. The methanol tolerance of wild-type and mutant PMLs as a function of methanol concentration is shown in Figure 5A. Due to activation of PML in low concentrations of methanol (not shown), the residual activity was normalized to activity remaining after incubation with

10% methanol. Similar activation of related lipases with short-chain alcohols has been demonstrated recently [30]. When the wild-type or Dieselzyme 1 is incubated for two hours at increasing methanol concentrations, both enzymes retain 50% activity in 55% methanol and are completely inactivated at 70% methanol. In contrast, Dieselzyme 2 retained ~20% residual activity after incubation with 70% methanol while the Dieselzyme 3 and Dieselzyme 4 retain greater than 70% of their activity after incubation in 70% methanol.

While we were able to improve the absolute methanol tolerance of PML during a relatively short incubation, the actual goal was to develop a mutant lipase that had better long-term resistance to the presence of methanol. In order to determine if the mutant lipases were also more resistant to methanol inactivation over a long period of time, wild-type or mutant PML was incubated for 24 hours in the presence of 50% methanol. Figure 5B shows the residual activity of wild-type and mutant PML from the different rounds of screening after treatment with 50% methanol as a function of time. Clearly the long-term resistance was greatly improved by the screening regimen, with Dieselzyme 4 retaining roughly 90% activity after incubation with 50% methanol for 24 hours. The *B. cepacia* lipase that is currently used industrially for biodiesel synthesis shows a comparable level of tolerance under these conditions. As shown in Figure 5C, however, when we raise the concentration of methanol to 70%, Dieselzyme 4 retains essentially full activity after 24 hrs, but the *B. cepacia* lipase retains only ~60%.

Other alcohols and water miscible solvents

To test whether the improvement in methanol tolerance also rendered Dieselzyme 4 more resistant to other water-miscible organic solvents, we incubated the enzymes in the presence of a variety of organic solvents at a concentration of 70% for 24 hours, and then assayed for residual hydrolysis activity (Figure 5C). In general, Dieselzyme 4 retained considerably more activity than the wild-type after incubation with organic solvents, with the exception of DMSO. DMSO surprisingly stimulates the wild-type PML enzyme, but has little effect on Dieselzyme 4.

We also compared the organic solvent tolerance of Dieselzyme 4 to the tolerance of the *B. cepacia* lipase. As shown in Figure 5C, Dieselzyme 4 is more tolerant than the *B. cepacia* lipase to the alcohols most commonly used in biodiesel synthesis, methanol and ethanol. The *B. cepacia* lipase is more tolerant to 1-propanol, acetonitrile and acetone, however, indicating that organic solvent tolerance is a complex property.

Transesterification and recycling of immobilized PML

As our goal is to develop a stable lipase that is less prone to methanol induced inactivation during biodiesel synthesis,

Table 1 List of mutations in engineered Dieselzymes

Construct	Mutations present
Dieselzyme 1	G181C/S238C*
Dieselzyme 2	G181C/S238C/K208N/L64I/A70T/F225L/Q277L
Dieselzyme 3	G181C/S238C/K208N/L64I/A70T/F225L/Q277L/ G202E/G266S/D270N/N175 [#]
Dieselzyme 4	G181C/S238C/K208N/L64I/A70T/F225L/Q277L/ G202E/G266S/D270N/N175/I255F/R33T

Mutations in bold indicate the new mutations that were identified in each round of directed evolution. The additive effect of mutations was analyzed by combining mutations through site-directed mutagenesis. Only the final mutant from each round is shown. * - Construct used as parent for the initial round of directed evolution. # - N175 mutation arose spontaneously while combining the G202E, G266S, and D270N mutations.

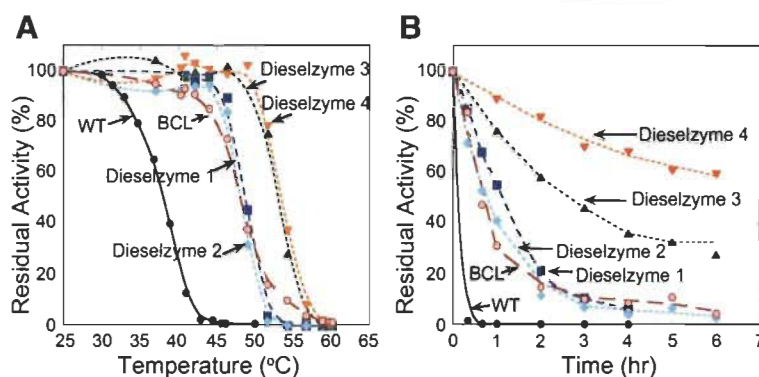


Figure 4 Temperature induced inactivation of Dieselzymes. (A) Inactivation as a function of temperature. Residual activity was measured after incubation for 1 hour at various temperatures. (B) Time course of thermal inactivation at 50°C. Results are the average of 3 independent experiments. Error-bars are omitted but are less than 5% in all cases. Lines are shown for clarity and represent the best-fit to the data points. WT PML (black circles), BCL (red open circles), Dieselzyme 1 (blue squares), Dieselzyme 2 (blue diamond), Dieselzyme 3 (green triangle), and Dieselzyme 4 (orange inverted triangle).

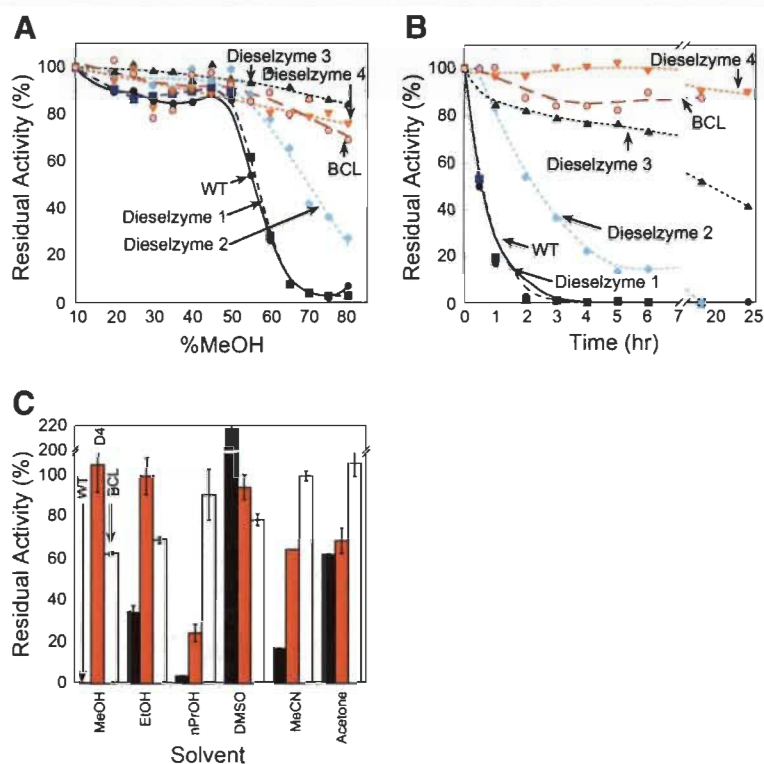
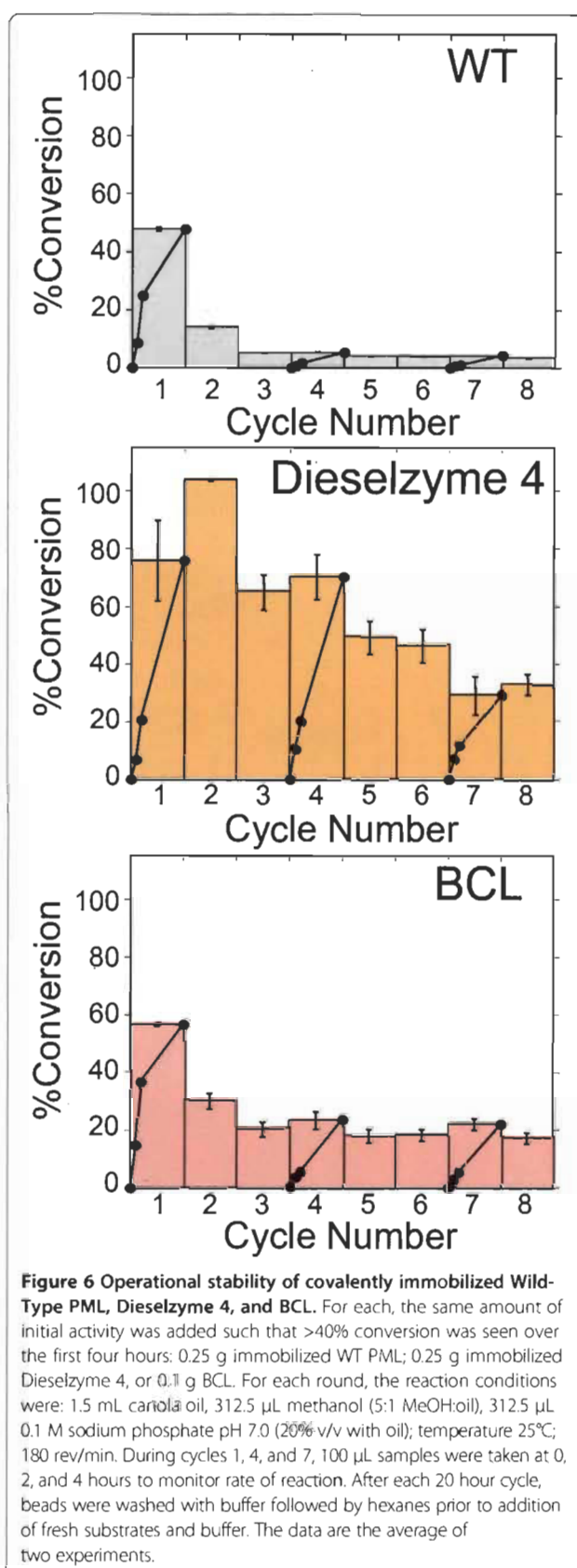


Figure 5 Methanol induced inactivation of Dieselzymes. (A) Methanol inactivation as a function of concentration. Samples were incubated for 2 hours and assayed after dilution to 10% methanol. Results are the average of 3 independent experiments. (B) Methanol inactivation as a function of time. Samples were incubated in 50% methanol for 24 hours. At indicated time points, samples were diluted to 10% methanol and assayed for residual activity. Error-bars are omitted but are less than 5% in all cases. Lines are shown for clarity only. Results are the average of 3 independent experiments. WT PML (black circles), BCL (red open circles), and Dieselzyme 1 (blue squares), Dieselzyme 2 (blue diamond), Dieselzyme 3 (green triangle), and Dieselzyme 4 (orange inverted triangle). (C) Inactivation profile of Wild-Type PML (black bar), Dieselzyme 4 (orange bar), and BCL (white bar) in response to various solvents. Enzymes were incubated for 24 hours in 70% solvent and diluted to 10% prior to assaying for residual activity. Results are the average of three independent experiments.

we tested the ability of Dieselzyme 4 to retain transesterification activity for many cycles over a long period of time. To monitor both transesterification activity and resistance to methanol induced inactivation during multiple rounds of synthesis, wild-type, Dieselzyme 4, and the industrial enzyme BCL were covalently immobilized onto hydrophobic oxirane functionalized beads. Covalent immobilization was used to ensure differences in activity were due to enzyme inactivation as opposed to loss by desorption. To monitor biodiesel synthesis, 200 μ g lipase covalently immobilized on 250 mg beads were added to a mixture of 0.625 ml of 50% methanol and 1.5 ml of canola oil. This mixture provides a \sim 5:1 molar ratio of methanol to triacylglycerol oil or a \sim 1.6 molar ratio of methanol per ester bond. The final water content was \sim 1:4 (or 26% w/w) of the oil. To test the response of the enzymes to prolonged exposure to methanol, a small amount of immobilized enzyme beads were added such that there was less than 10% conversion in the first hour. A similarly low extent of conversion for all enzymes during the first few hours was used to ensure differences seen between enzymes were due to the time of exposure to methanol. For each cycle the progress of the reaction was monitored after 20 hours. For the first, fourth and seventh cycles, we also monitored the reaction progress at 2, 4 and 6 hours to observe initial rates. To restart the reaction at each cycle, the beads were separated by filtration and washed prior to reuse in a subsequent round of biodiesel synthesis. A water wash was performed to remove glycerol and residual methanol followed by a wash with hexanes to remove residual oil and fatty acid methyl esters.

The results of multiple rounds of transesterification can be seen in Figure 6. During the first cycle the initial rate of biodiesel synthesis was linear and showed a similar rate of conversion for all three immobilized enzymes, with 24.7%, 20.6%, and 36.5% conversion to methyl esters for WT, Dieselzyme 4, and BCL respectively after 4 hours. After 4 hours the rate of synthesis decreased for both WT and BCL, such that only \sim 47.7% and \sim 56.4% total conversion was seen after 20 hours. In contrast, Dieselzyme 4 continued at a high rate of synthesis, converting \sim 76% of the canola oil to biodiesel within 20 hours. It is unclear why conversion by Dieselzyme 4 did not reach 100% in the first round. However, it is obvious that even though there was higher initial BCL activity, Dieselzyme 4 was able to convert more oil over the longer incubation.

The improved methanol tolerance of Dieselzyme 4 becomes even more apparent when the immobilized enzymes are reused in a second cycle. After a single reuse, the wild-type PML lost nearly all of its ability to catalyze transesterification with less than 15% of the canola oil being converted to biodiesel in the second round.



Similarly, BCL also showed a decreased yield of biodiesel with only 30% conversion of TAG to FAME in the second round. In contrast, Dieselzyme 4 retained complete activity in the second round. Overall, the rate of inactivation of Dieselzyme 4 was significantly slower than that of BCL as Dieselzyme 4 retained nearly the same amount of activity after round 4 compared to round 1 and still converted 50% of canola oil to biodiesel after the fifth cycle (compared to only 18% for BCL after cycle 5). These results show that not only is Dieselzyme 4 active for transesterification, but Dieselzyme 4 is significantly more resistant to methanol induced inactivation during biodiesel synthesis compared to wild-type PML and even outperforms BCL.

Productivity of Dieselzyme 4

Fjerbaek et al. define productivity in terms of kg of biodiesel produced per kg of catalyst [9]. Their cost analysis suggests that a 23-fold improvement in lipase productivity from the current high water mark of 7400 kg/kg to ~170,000 kg/kg would make enzyme catalyzed biodiesel production cost neutral with the base catalyzed reaction (not including the many advantages of lipase catalysis). To measure the potential productivity of Dieselzyme 4, we used milder conditions than the recycling test above, while still employing single additions of methanol. 200 µg of purified enzyme was immobilized on 0.5 g of oxirane functionalized beads and used to convert 5 mL (4.5 g) of refined oil to biodiesel using 2 mL of 40% methanol (equivalent to 4:1 molar ratio of methanol:oil and 25% water (v/v water/oil)). As shown in Additional file 3: Figure S3, greater than 50% of the input oil was converted to biodiesel in the first cycle. Similar to the recycling experiments above, Dieselzyme 4 retains ~50% of the initial activity after 6 cycles (1.1 g biodiesel produced in cycle 5 compared to 2.4 g biodiesel produced in cycle 1). Overall, the productivity of immobilized Dieselzyme 4 is ~46,000 kg/kg after 6 cycles. Assuming an exponential decay of activity, the productivity of Dieselzyme 4 is projected to plateau at ~82,000 kg/kg (Additional file 3: Figure S3). Thus, the productivity of Dieselzyme 4 is between 46,000 and 82,000 kg/kg, which is within a 2- to 4-fold improvement in activity, longevity, process enhancements or enzyme production costs of being cost equivalent with base catalysis.

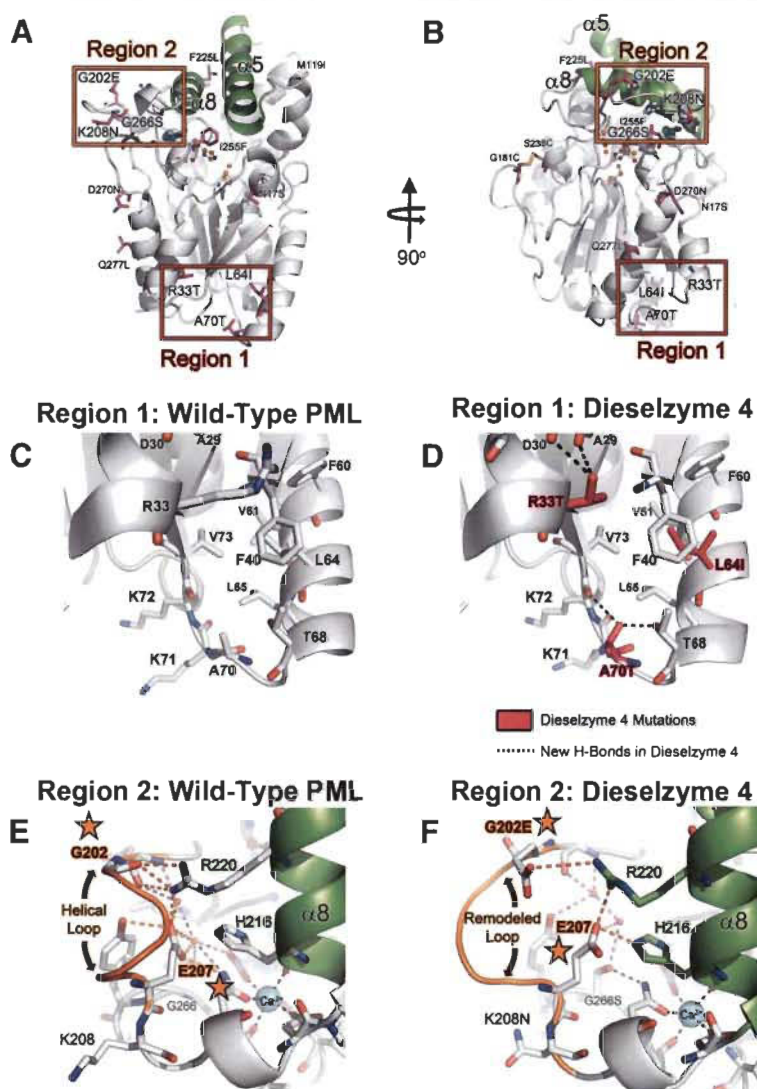
Structural basis for evolved methanol tolerance

The structure of Dieselzyme 4 was determined by x-ray crystallography to examine the effects of the mutations on the PML structure. Besides the introduced disulfide bond discussed above, 11 new mutations arose from in vitro evolution. An analysis of the crystal structure of Dieselzyme 4 suggests that 6 mutations cluster into two distinct areas, (Region 1 and Region 2) that may be

important for the improvement in methanol tolerance (Figure 7A and 7B). The other 5 mutations may be neutral with respect to methanol tolerance because they are located on the surface (F225L, D270N, Q277L) or in regions that do not lead to obvious differences (N17S, I255F).

Region 1 comprises the mutations R33T, L64I, and A70T that cluster on or near helices $\alpha 1$ and $\alpha 2$ at opposite end of the protein from the lid region (Figure 7C and 7D). Well-defined electron density was seen for all three of the new side chains in this region (Additional file 2: Figure S2B). Mutation of leucine to isoleucine at position 64 provides more interactions with side-chains F40 and F60, possibly improving packing to stabilize helix $\alpha 2$. Distance measurements of the A70T mutation, which arose in the first round of error-prone PCR (Dieselzyme 2), shows that the hydroxyl from the threonine forms a new H-bond interaction with T68 and the backbone carbonyl of K72 (Figure 7D). Similarly, the R33T mutation that arose after the final round of error-prone PCR (see Dieselzyme 4) forms new H-bond interactions between the threonine hydroxyl and the backbone carbonyls from A29 and D30. In polar organic solvents such as methanol, it has been hypothesized that inactivation and unfolding of proteins is due to a combination of stripping of water from the protein surface and hydrophobic collapse of buried non-polar regions due to the destabilizing effect of aliphatic alcohols on tertiary interactions [31,32]. It is possible that the L64I mutation provides improved packing interactions that counteract the hydrophobic collapse of Region 1 in response to methanol while the R33T and A70T provide stabilizing H-bonds. Due to the lower dielectric constant of methanol, these introduced H-bonds may be further strengthened [33]. It has also been reported for CALB lipase [34], phospholipase A1 [35], and metalloprotease [36], that the introduction of polar residues can lead to increases in solvent tolerance.

Region 2 mutations (Figure 7E and 7F) cluster near the Ca^{2+} binding site, and cause an unexpected remodeling of the loop between residues 200 and 208. The introduction of polar side chains may also stabilize the Ca^{2+} binding site and provide additional polar interactions between the loop and lid helix $\alpha 8$ that may resist the destabilization due to methanol. While the density of the remodeled loop is not as well-defined as the rest of the protein (Additional file 2: Figure S2C), there is sufficient electron density to provide a possible rationale for the observed improvement in methanol tolerance. Loop mutations G202E, K208N, and G266S and are located close to the Ca^{2+} binding site which is essential for lipase stability and activity. The most striking change in the loop is a direct result of the G202E mutation. The introduced glutamate side chain at position 202 is positioned ~4 Å from R220 in an orientation that could



form a new salt-bridge which may provide stability in the low dielectric methanol environment. In order to make this potential new salt-bridge, the helical region between residues 203 and 207 seen in wild-type PML partially unravels and may increase flexibility of the loop region to allow new interactions to occur. For instance, the increased flexibility due to unraveling of the helix between 203 and 207 may allow E207 to move and form a new salt-bridge with R220 and

conserved H216. However, the electron density surrounding R220 is split, suggesting that R220 is in at least two conformations: one that can interact with E207 and one that potentially interacts with the introduced G202E mutation (Figure 7F). Additionally, while the K208N mutation is well defined and does not alter the backbone or side chain conformation at position 208, it is possible that removal of the charged Lys also increases the flexibility of the loop region, facilitating

the new interactions made by the G202E mutation and E207 that stabilize this region in the presence of methanol.

Interestingly, in addition to the polar residue insertion (K208N/G202E) and loop remodeling described above, the G266S mutation forms another new polar interaction, albeit with a crystallographic water and the side chain N210 that directly coordinates the Ca^{2+} . It has been proposed that a major effect of alcohols on proteins is the stripping of water from the protein surface leading to destabilization [37]. Previous experiments with the homologous lipases from *B. glumae* have also indicated that the bound Ca^{2+} plays an important role in lipase stability [38]. The introduction of a new polar interaction from G266S may help keep water bound and further stabilize the bound Ca^{2+} in Dieselzyme 4 by providing a new H-bond to N210. Together, the polar mutations in Region 2 appear to provide stabilizing interactions with the lid helix $\alpha 8$ (through R220 and H216) and to potentially stabilize the Ca^{2+} binding site that may be sensitive to the destabilizing effects of methanol and other alcohols.

Conclusion

We have employed directed evolution for the first time to improve properties of a lipase for biodiesel production. Unlike the *B. cepacia* lipase, PML is highly expressed in an active form in *E. coli*, making it a viable platform for engineering efforts. The evolved PML enzyme, Dieselzyme 4, catalyzes efficient synthesis of biodiesel even in the presence of a high concentration of water and has better methanol tolerance and heat stability than the top industrially used lipase from *B. cepacia*. The success of the screening protocol and identification of regions on the enzyme central to improved stability makes Dieselzyme 4 a viable platform for further engineering by further directed evolution, perhaps by targeted methods such as CAST [39] or ISM [40] to develop better catalysts for biodiesel production in the future.

Methods

Materials

All chemicals were of analytical grade or better. 4-nitrophenyl palmitate, and Triton X100 were from Sigma. 1-naphthyl palmitate and Fast Blue B were from MPI Biochemicals. All other solvents were from Fluka. Amano lipase PS (*Burkholderia cepacia*) was purchased from Sigma and was purified before use using a HiTrapQ column (GE life sciences) and dialyzed into 20 mM Tris-Cl pH 7.5, 0.1 M NaCl. Primers for cloning and mutagenesis were ordered from Valuegene. Refined canola oil used for transesterification was from the local market.

Construction of mutants by site-directed mutagenesis

Site-directed mutagenesis was performed using the Quick-change Site Directed Mutagenesis Kit (Stratagene) according to the manufacturer's directions. For generation of double mutants, a modified megaprimer method was used [41]. Briefly, in an initial PCR, a forward and reverse mutagenic primer, each containing a different mutation, was used to amplify a short segment of the PML gene using Taq Hot-Start Supermix (Denville) to generate a megaprimer. In a second PCR reaction, 5 μL of the initial PCR was mixed with 30 ng plasmid template, 0.2 μM dNTP, Pfu HotStart II Reaction Buffer, and 2.5 U of Pfu HotStart Fusion II DNA Polymerase (Stratagene) in a 50 μL reaction. The reaction was cycled as for the QuikChange reaction. The resulting PCR product was digested with *DpnI* and used to transform BL21Gold(DE3) or XL10Gold directly.

Construction of PML mutant library by error-prone PCR

The wild-type lipase gene from *Proteus mirabilis* was cloned into a pET28a vector as previously described and used as template for error-prone PCR. Random mutagenesis was performed using the Genemorph II Kit (Stratagene) according to the manufacturer's instructions to ensure a 1-2% error rate per gene (1-5 amino acid changes). Primers which flank the 5'*NdeI* and 3'*EagI* restriction sites were used for amplification. An appropriate amount of DNA template was used to generate between 1 and 5 mutations per 1 kb. Briefly, 0.1 ng template was mixed with 0.2 mM each dNTP, 0.2 μM each primer, 1X Mutazyme II Buffer, and 2.5 U Mutazyme II in 50 μL . The PCR was incubated at 95°C for 2 minutes followed by 45 cycles of 95°C for 30 s, 58°C for 30 s, 72°C for 1 min, with a final extension at 72°C for 10 minutes. The resulting PCR product was purified using a QIAquick spin column (Qiagen) and digested overnight with *NdeI* and *EagI* (NEB) at 37°C. The digested product was gel purified and ligated into *NdeI/EagI* digested pET28 using T4 DNA ligase (NEB) for 16 hours at 16°C. The resulting library was directly transformed into chemically competent BL21Gold(DE3) and plated on LB-agar containing 50 $\mu\text{g}/\text{mL}$ kanamycin for screening and analysis. To confirm the desired 1-2% error rate per gene, plasmid was isolated from 20 colonies and sequenced (Genewiz).

Library screening for MeOH tolerance

Screening for improved methanol tolerance was accomplished by a colony lift screening protocol with a PML mutant library expressed in BL21Gold(DE3). Following overnight incubation at 37°C, transformants (~600/plate) were lifted onto sterile filter circles (Whatman 410) and placed colony-up on a LB-agar plate containing 50 $\mu\text{g}/\text{mL}$ kanamycin and 1 mM IPTG

for 2–3 hours at 18°C to induce protein expression. The filter was then immersed in lysis solution (50 mM sodium phosphate pH 7.5, 0.1 M NaCl, 0.1% Triton X100, 1 mg/mL lysozyme) at 25°C for 1 hour. Lysis solution was decanted and replaced by MeOH solution containing 0.1% Triton X100 and incubated at a selected temperature for a desired amount of time. After incubation, the MeOH solution was decanted and the filters were developed by overlaying with 1 mM 1-naphthyl palmitate, 3 mM Fast Blue B, 0.5% Triton X100 dispersed in 0.5% agar. After 10 minutes, mutants displaying residual activity were identified by formation of a purple color due to the azo dye formed between 1-naphthol and Fast Blue B (Figure 3). The corresponding colony was then isolated from the master plate for validation and further characterization.

For validation, positives from the filter screen were grown in 10 mL LB containing 50 µg/mL kanamycin to OD₆₀₀ of 0.6 and protein expression was induced with 0.5 mM IPTG for 16 hours at 18°C. The cells were pelleted (6K rcf × 30 min; Sorvall GS-3 rotor), resuspended in buffer (50 mM sodium phosphate pH 7.5, 0.1 M NaCl, 5 mM imidazole), lysed by sonication (5 × 30 s pulse) and clarified by centrifugation (20K rcf × 30 min; Sorvall SS-34 rotor). To validate, 50 µL of the supernatant was incubated with 70% MeOH for 1 hour and then diluted to 10% MeOH prior to being assayed with 1 mM *p*-nitrophenyl palmitate (pNPP). Residual activity was defined as the activity compared to a sample incubated at 10% MeOH for 1 hour. Mutations were confirmed by sequencing (Genewiz) and the beneficial mutations were combined by site-directed mutagenesis and reassayed for improved MeOH tolerance. The best performing combined mutants were then used as parents for subsequent rounds of directed evolution.

Expression and purification of WT and mutant PMLs

Over-expression of wild-type and mutant PML was carried out in *E. coli* BL21Gold(DE3) (Agilent). Single transformants were transferred to 2 L of Luria-Bertani (LB) media containing kanamycin for incubation at 37°C and grown to an OD₆₀₀ of 0.6. Protein expression was induced with 0.5 mM IPTG at 16°C for 16 hours. Cells were harvested by centrifugation and purified by Ni-NTA chromatography (Qiagen) as described previously. Wild-type and mutant PMLs were then dialyzed into 20 mM Tris-HCl pH 7.5 containing 100 mM NaCl and flash frozen as droplets in liquid N₂ prior to storage at -80°C.

Enzyme kinetics and characterization

Kinetic assays were carried out in 96 well microtiter plates (Grenier Bio-One) with purified protein. For kinetic analysis a stock solution of 86.13 mM pNPP was prepared in 1:1 Acetonitrile:Triton X100. The initial rate

of conversion of pNPP to *p*-nitrophenol was monitored at 405 nm over the first minute using a plate reader (SpectraMax M5, Molecular Devices). For temperature or methanol incubation studies, the residual activity was assayed using 1 mM pNPP in 50 mM phosphate pH 7.0. For temperature inactivation, 60 µL of 150 nM enzyme was incubated in 50 mM phosphate pH 7.0 in thin walled PCR using a PCR cyclor (Mastercycler ProS, Eppendorf). The residual activity remaining as a function of incubation for 1 hour at various temperatures or as a function of time at 50°C was determined by hydrolysis of pNPP. Activity was normalized to activity at 25°C. For methanol tolerance, 1.5 µM enzyme was incubated with 50 mM phosphate pH 7.0 containing various concentrations of methanol. After 2 hours, the enzyme was diluted 10-fold so that the final methanol concentration was 10% and the final enzyme concentration was 150 nM. For inactivation by methanol over time, the enzyme was incubated in 50% methanol at 25°C. At various time points, 10 µL aliquots were diluted 1:10 with 90 µL 5.5% methanol (10% final) and assayed for residual activity compared to incubation with 10% methanol as described above.

Immobilization and transesterification

Purified *B. cepacia*, PML, and mutant lipases (Dieselzymes) were covalently immobilized on hydrophobic Immobead 350 oxirane functionalized beads (ChiralVision) prior to use. Beads (0.25 g) were washed once with 10 mL methanol followed by two washes with 10 mL 0.1 M sodium phosphate pH 7.0. The buffer was decanted and 1 mL enzyme at 0.2 mg/mL in 20 mM Tris-HCl pH 7.5, 0.1 M NaCl was added. Immobilization was allowed to proceed for 16–20 hours at 25°C. All 3 enzymes studied were immobilized to a similar degree (≥95%) as monitored by OD_{280nm} and by residual activity remaining in the supernatant.

After immobilization, the buffer was decanted and the beads were used for transesterification without further modification. For the synthesis reaction, 0.625 mL of 50% methanol (1:1 methanol:0.1 M phosphate pH 7.0) was added to the beads followed by 1.5 mL refined canola oil. The solution was gently mixed with a vortex and then placed on a shaker at 200 rpm at 25°C. An initial reaction was performed with 0.1 g beads to establish the transesterification rate for each construct. For the recycling experiment, an appropriate amount of beads (0.25 g, 0.25 g, and 0.08 g of wild-type PML, Dieselzyme 4, and BCL respectively) were added such that ~5–10% conversion was reached in the first hour. After each 20 hour cycle, 10 µL aliquots of the oil layer were taken for analysis by gas chromatography (below). To monitor the effect of reuse on initial rate, 10 µL were taken at 2 and 4 hours during cycles 1, 4,

and 7. For reuse, the beads were recovered by filtration, washed with 5 mL buffer followed by 5 mL hexanes and then allowed to dry before adding fresh methanol, buffer, and oil.

Quantification of fatty acid methyl esters

The extent of transesterification was monitored by gas chromatography (GC). For analysis, 10 μ L samples were diluted with 1 mL hexane spiked with 0.5 mg/mL methyl heptadecanoate (internal standard). 1 μ L of the samples at a split ratio of 50:1 were analyzed on an Agilent 5890 Series II GC with flame ionized detector using an HP-INNOWax column (0.25 mm \times 30 m, Agilent). The carrier gas was helium with a flow rate of 5 mL/min. The oven temperature was kept at 200°C for 3 min and then raised to 230 at 5°C/min then to 250 at 20°C/min and held at 250°C for 9 min. The injector and detector temperatures were kept at 230 and 330°C respectively. The percent conversion was determined by comparison to a biodiesel sample prepared from refined canola oil using a large excess of free *Burkholderia cepacia* lipase at a 5:1 MeOH:oil ratio in the presence of 5% water as described previously [23].

Crystallization, structure determination, and refinement of mutant PML

Crystallization trials for Dieselzyme 4 were performed using purified His-tagged lipase with no additives. Drops were generated by mixing 2 μ L purified protein at 9 mg/mL with 2 μ L well solution. Large crystals of the mutant PML formed in many conditions between one day and two weeks. One condition, Qiagen PACT condition #38 (1X MMT pH 5, 20% PEG 1500), was optimized and gave large crystals (0.2 \times 0.2 \times 0.2 mm) within one week. Prior to data collection, crystals were soaked in crystallization well solution plus 15% glycerol and flash frozen in liquid nitrogen. Data were collected in house on a Rigaku FRE + x-ray generator equipped with an ADSC Quantum 4 CCD detector at 100 K. Diffraction images were indexed, integrated, and scaled with Denzo and Scalepack. Initial phases were determined by Molecular Replacement using PHASER in CCP4i. The wild-type PML (PDBID 4GW3) was used as the search model. The resulting model was further refined via iterative rounds of model building and refinement in COOT and Refmac5 over the resolution range 50–1.8 Å. The final model was deposited in the Protein Data Bank with PDBID 4HS9. Data statistics can be found in Additional file 4: Table S1.

Additional files

Additional file 1: Figure S1. Thermal inactivation of Dieselzyme 1 by incubation at 50°C as a function of time. Results shown are an average of 3 independent experiments. Error-bars are omitted but are less than 5%

in all cases. Lines are shown for clarity and represent the best-fit to the data points: WT PML (black circles), BCL (red open circles), and Dieselzyme 1 (blue squares).

Additional file 2: Figure S2. Electron density of the Dieselzyme 4 crystal structure in the vicinity of (A) The introduced disulfide bond as a result of the G181C/S238C mutation (B) Region 1 and (C) Region 2 mutations. The 2Fo-Fc map is shown contoured to 1 σ .

Additional file 3: Figure S3. Productivity of Dieselzyme 4. The enzyme was covalently immobilized on oxirane beads. The amount (g) of product produced in each round is plotted versus cycle number (black circle). The black line represents an exponential fit assuming a first order decay. The cumulative amount of biodiesel produced at the end of each round is also shown (red diamonds). The projected cumulative total biodiesel produced, estimated from the exponential decay, is shown as a red dashed line. Each reaction consisted of 200 μ g purified enzyme immobilized on 0.5 g beads, 4.5 g canola oil, 2 mL 40% aqueous methanol. The reactions were incubated for 24 hours at 25°C on a rotary shaker at 200 rpm. The reactions were performed in duplicate and the results quantified by gas chromatography.

Additional file 4: Table S1. Crystallization, data collection and refinement statistics.

Abbreviations

FAME: Fatty acid methyl ester; TAG: Triacylglycerol; MeOH: Methanol; EtOH: Ethanol; FFA: Free fatty acids; DMSO: Dimethylsulfoxide; DMF: Dimethylformamide; MeCN: Acetonitrile; PML: *Proteus mirabilis* lipase; BCL: *Burkholderia cepacia* lipase; CALB: *Candida Antarctica* lipase B; PCR: Polymerase chain reaction; LB: Luria-bertani media; IPTG: Isopropyl- β -D-1-thiogalactopyranoside; pNPP: *Para*-nitrophenyl palmitate; CAST: Combinatorial active-site saturation test; ISM: Iterative saturation mutagenesis.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

TPK designed experiments, carried out the biochemical experiments, solved the crystal structures, and drafted the manuscript. BS, DMC, and GLH carried out screening and mutagenesis experiments. MB participated in the analysis of results. JUB conceived of the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by grants from the US Department of Energy and from the Camille and Henry Dreyfus Foundation. D.M.C. and M.B. acknowledge the support of a PROTEO scholarship and NSERC grant (Discovery 138654).

Author details

¹Department of Chemistry and Biochemistry, UCLA-DOE Institute of Genomics and Proteomics, Molecular Biology Institute, University of California, Los Angeles, USA. ²Département de chimie-biologie, Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada.

Received: 4 December 2012 Accepted: 1 May 2013

Published: 7 May 2013

References

1. Luque R, Lovett JC, Datta B, Clancy J, Campelo JM, Romero AA: **Biodiesel as feasible petrol fuel replacement: a multidisciplinary overview.** *Energy Environ Sci* 2010, **3**:1706–1721.
2. Robles-Medina A, Gonzalez-Moreno PA, Esteban-Cerdan L, Molina-Grima E: **Biocatalysis: towards ever greener biodiesel production.** *Biotechnol Adv* 2009, **27**:398–408.
3. Szczesna Antczak M, Kubiak A, Antczak T, Bielecki S: **Enzymatic biodiesel synthesis - key factors affecting efficiency of the process.** *Renew Energy* 2009, **34**:1185–1194.

4. Jaeger K-E, Eggert T: **Lipases for biotechnology.** *Curr Opin Biotechnol* 2002, **13**:390-397.
5. Hasan F, Shah AA, Hameed A: **Industrial applications of microbial lipases.** *Enzyme Microb Technol* 2006, **39**:235-251.
6. Parawira W: **Biotechnological production of biodiesel fuel using biocatalysed transesterification: A review.** *Crit Rev Biotechnol* 2009, **29**:82-93.
7. Sotoft LF, Rong B-G, Christensen KV, Norddahl B: **Process simulation and economical evaluation of enzymatic biodiesel production plant.** *Bioresour Technol* 2010, **101**:5266-5274.
8. Biermann U, Bornscheuer U, Meier MAR, Metzger JO, Schäfer HJ: **Oils and fats as renewable raw materials in chemistry.** *Angew Chem Int Ed* 2011, **50**:3854-3871.
9. Fjerbaek L, Christensen KV, Norddahl B: **A review of the current state of biodiesel production using enzymatic transesterification.** *Biotechnol Bioeng* 2009, **102**:1298-1315.
10. Kourist R, Brundiek H, Bornscheuer UT: **Protein engineering and discovery of lipases.** *Eur J Lipid Sci Technol* 2010, **112**:64-74.
11. Tan T, Lu J, Nie K, Deng L, Wang F: **Biodiesel production with immobilized lipase: a review.** *Biotechnol Adv* 2010, **28**:628-634.
12. Joshi GK, Vinay S: **Bacterial lipases: classification, properties and applications in biotechnology.** *Res J Biotechnol* 2007, **2**:50-56.
13. Bornscheuer UT, Huisman GW, Kazlauskas RJ, Lutz S, Moore JC, Robins K: **Engineering the third wave of biocatalysis.** *Nature* 2012, **485**:185-194.
14. Moore JC, Arnold FH: **Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents.** *Nat Biotechnol* 1996, **14**:458-467.
15. You L, Arnold FH: **Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide.** *Protein Eng* 1996, **9**:77-83.
16. Song JK, Rhee JS: **Enhancement of stability and activity of phospholipase A1 in organic solvents by directed evolution.** *Biochim Biophys Acta-Protein Struct Molec Enzym* 2001, **1547**:370-378.
17. Fang Y, Lu Y, Lv F, Bie X, Zhao H, Wang Y, Lu Z: **Improvement of alkaline lipase from *Proteus vulgaris* T6 by directed evolution.** *Enzyme Microb Technol* 2009, **44**:84-88.
18. Reetz MT, Soni P, Fernandez L, Gumulya Y, Carballeira JD: **Increasing the stability of an enzyme toward hostile organic solvents by directed evolution based on iterative saturation mutagenesis using the B-FIT method.** *Chem Commun* 2010, **46**:8657-8658.
19. Valero F: **Heterologous Expression Systems for Lipases: A Review.** In *Lipases and Phospholipases Volume 861*. Edited by Sandoval G, Walker JM. New York: Humana Press; 2012. **861**:161-178. *Methods in Molecular Biology*.
20. Gao B, Su E, Lin J, Jiang Z, Ma Y, Wei D: **Development of recombinant *Escherichia coli* whole-cell biocatalyst expressing a novel alkaline lipase-coding gene from *Proteus* sp. for biodiesel production.** *J Biotechnol* 2009, **139**:169-175.
21. Arpigny JL, Jaeger KE: **Bacterial lipolytic enzymes: classification and properties.** *Biochem J* 1999, **343**(Pt 1):177-183.
22. Connor MR, Liao JC: **Microbial production of advanced transportation fuels in non-natural hosts.** *Curr Opin Biotechnol* 2009, **20**:307-315.
23. Kaieda M, Samukawa T, Kondo A, Fukuda H: **Effect of methanol and water contents on production of biodiesel fuel from plant oil catalyzed by various lipases in a solvent-free system.** *J Biosci Bioeng* 2001, **91**:12-15.
24. Le QAT, Joo JC, Yoo YJ, Kim YH: **Development of thermostable *Candida antarctica* lipase B through novel in silico design of disulfide bridge.** *Biotechnol Bioeng* 2012, **109**:867-876.
25. Han Z, Han S, Zheng S, Lin Y: **Enhancing thermostability of a *Rhizomucor miehei* lipase by engineering a disulfide bond and displaying on the yeast cell surface.** *Appl Microbiol Biotechnol* 2009, **85**:117-126.
26. Nardini M, Lang DA, Liebeton K: **Crystal structure of *Pseudomonas aeruginosa* lipase in the open conformation THE prototype for family I.1 of bacterial lipases.** *J Biol Chem* 2000, **275**:31219-31225.
27. Kim KK, Song HK, Shin DH, Hwang KY, Suh SW: **The crystal structure of a triacylglycerol lipase from *Pseudomonas cepacia* reveals a highly open conformation in the absence of a bound inhibitor.** *Structure* 1997, **5**:173-185.
28. Gatti-Lafranconi P, Natalello A, Rehm S, Doglia SM, Pleiss J, Lotti M: **Evolution of stability in a cold-active enzyme elicits specificity relaxation and highlights substrate-related effects on temperature adaptation.** *J Mol Biol* 2010, **395**:155-166.
29. Xu YY, Du W, Zeng J, Liu DH: **Conversion of soybean oil to biodiesel fuel using lipozyme TL IM in a solvent-free medium.** *Biocatal Biotransform.* 2004, **22**:45-48.
30. Whangsuk W, Sungkeeree P, Thiengmag S, Kerdwong J, Sallabhan R, Mongkolsuk S, Loprasert S: **Gene cloning and characterization of a novel highly organic solvent tolerant lipase from *Proteus* sp. SW1 And its application for biodiesel production.** *Mol Biotechnol* 2013, **53**:55-62.
31. Sashi P, Yasin UM, Bhuyan AK: **Unfolding action of alcohols on a highly negatively charged state of cytochrome c.** *Biochemistry* 2012, **51**:3273-3283.
32. Tanaka A: **Differential scanning calorimetric studies on the thermal unfolding of *Pseudomonas cepacia* lipase in the absence and presence of alcohols.** *J Biochem* 1998, **123**:289-293.
33. Hwang S, Shao Q, Williams H, Hilty C, Gao YQ: **Methanol strengthens hydrogen bonds and weakens hydrophobic interactions in proteins - a combined molecular dynamics and NMR study.** *J Phys Chem B* 2011, **115**:6653-6660.
34. Park H, Joo J, Park K, Yoo Y: **Stabilization of *Candida antarctica* lipase B in hydrophilic organic solvent by rational design of hydrogen bond.** *Biotechnol Bioprocess Eng* 2012, **17**:722-728.
35. Song JK, Rhee JS: **Enhancement of stability and activity of phospholipase A1 in organic solvents by directed evolution.** *Biochim Biophys Acta Protein Struct Mol Enzymol* 2001, **1547**:370-378.
36. Badoei-Dallard A, Khajeh K, Asghari SM, Ranjbar B, Karbalaee-Heidari HR: **Enhanced activity and stability in the presence of organic solvents by increased active site polarity and stabilization of a surface loop in a metalloprotease.** *J Biochem* 2010, **148**:231-238.
37. Gorman LAS, Dordick JS: **Organic solvents strip water off enzymes.** *Biotechnol Bioeng* 1992, **39**:392-397.
38. Invernizzi G, Papaleo E, Grandori R, De Gioia L, Lotti M: **Relevance of metal ions for lipase stability: structural rearrangements induced in the *Burkholderia glumae* lipase by calcium depletion.** *J Struct Biol* 2009, **168**:562-570.
39. Reetz MT, Bocla M, Carballeira JD, Zha D, Vogel A: **Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test.** *Angew Chem* 2005, **117**:4264-4268.
40. Reetz MT, Carballeira JD: **Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes.** *Nat Protoc* 2007, **2**:891-903.
41. Barik S: **Megaprimer PCR.** In *PCR Cloning Protocols Volume 192*. Edited by Chen B-Y, Janes HW. New Jersey: Humana Press; 2002. **192**:189-196. *Methods in Molecular Biology*.

doi:10.1186/1754-6834-6-70

Cite this article as: Korman *et al.*: **Dieselzymes: development of a stable and methanol tolerant lipase for biodiesel production by directed evolution.** *Biotechnology for Biofuels* 2013 **6**:70.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



A.4 Conclusion générale

Les déterminants structuraux impliqués dans la thermostabilité de deux nouvelles enzymes lipolytiques bactériennes, une carboxylesterase thermostable de *G. thermodenitrificans*, EstGtA2, et une lipase psychrostable de *P. mirabilis*, PML, ont été étudiés par deux approches différentes. Une approche rationnelle a été utilisée pour identifier certains éléments structuraux conservés contrôlant la thermostabilité de EstGtA2 alors qu'une approche d'évolution accélérée a été utilisée pour accroître la thermostabilité de PML et de comprendre l'origine de sa nouvelle stabilité. La comparaison des résultats issus de ces deux approches a révélé d'étonnantes similitudes entre les éléments structuraux responsables de la thermostabilité. Dans les deux cas, les interactions électrostatiques se sont vues jouer un rôle essentiel dans la thermostabilité de ces enzymes lipolytiques. Un groupe de ponts salins conservés s'est révélé être essentiel pour la thermostabilité de EstGtA2 alors que des nouvelles interactions polaires, principalement de nouveaux ponts hydrogènes et ponts salins sont responsables de la thermostabilisation de PML. Ce constat intéressant suggère que les interactions électrostatiques jouent un rôle essentiel à haute température et que celles-ci sont particulièrement stabilisantes pour les enzymes lipolytiques qui opèrent dans des milieux de faibles constantes diélectriques. Dans des conditions où les interactions hydrophobes perdent de leurs effets stabilisants, les interactions électrostatiques deviennent quant à elles particulièrement importantes.

ANNEXE B

FIGURES SUPPLÉMENTAIRES

		Second Position				
	U	C	A	G		
U	UUU } Phe	UCU } Ser	UAU } Tyr	UGU } Cys	U	
	UUC }	UCC }	UAC }	UGC }	C	
	UUA } Leu	UCA }	UAA Stop	UGA Stop	A	
	UUG }	UCG }	UAG Stop	UGG Trp	G	
C	CUU } Leu	CCU } Pro	CAU } His	CGU } Arg	U	
	CUC }	CCC }	CAC }	CGC }	C	
	CUA }	CCA }	CAA } Gln	CGA }	A	
	CUG }	CCG }	CAG }	CGG }	G	
A	AUU } Ile	ACU } Thr	AAU } Asn	AGU } Ser	U	
	AUC }	ACC }	AAC }	AGC }	C	
	AUA }	ACA }	AAA } Lys	AGA } Arg	A	
	AUG Met	ACG }	AAG }	AGG }	G	
G	GUU } Val	GCU } Ala	GAU } Asp	GGU } Gly	U	
	GUC }	GCC }	GAC }	GGC }	C	
	GUA }	GCA }	GAA } Glu	GGA }	A	
	GUG }	GCG }	GAG }	GGG }	G	

Figure B.1 Table des codons (code génétique).

Correspondance entre les acides nucléiques (codon de l'ARNm) et les acides aminés telle qu'établie par Francis Crick en 1966.

Table 1

Enzymes used in various industrial segments and their applications.

Industry	Enzyme class	Application
Detergent (laundry and dish wash)	Protease → Amylase Lipase Cellulase Mannanase	Protein stain removal Starch stain removal Lipid stain removal Cleaning, color clarification, anti-redeposition (cotton) Mannan stain removal (reappearing stains)
Starch and fuel	Amylase Amyloglucosidase Pululanase Glucose isomerase Cyclodextrin-glycosyltransferase Xylanase	Starch liquefaction and saccharification Saccharification Saccharification Glucose to fructose conversion Cyclodextrin production Viscosity reduction (fuel and starch)
Food (including dairy)	Protease → Lipase Lactase Pectin methyl esterase Pectinase Transglutaminase	Protease (yeast nutrition – fuel) Milk clotting, infant formulas (low allergenic), flavor Cheese flavor Lactose removal (milk) Firming fruit-based products Fruit-based products Modify visco-elastic properties
Baking	Amylase Xylanase → Lipase Phospholipase Glucose oxidase Lipoxygenase Protease Transglutaminase	Bread softness and volume, flour adjustment Dough conditioning Dough stability and conditioning (<i>in situ</i> emulsifier) Dough stability and conditioning (<i>in situ</i> emulsifier) Dough strengthening Dough strengthening, bread whitening Biscuits, cookies Laminated dough strengths
Animal feed	Phytase Xylanase β-Glucanase	Phytate digestibility – phosphorus release Digestibility Digestibility
Beverage	Pectinase Amylase β-Glucanase Acetolactate decarboxylase	De-pectinization, mashing Juice treatment, low calorie beer Mashing Maturation (beer)
Textile	Laccase Cellulase Amylase Pectate lyase Catalase Laccase Peroxidase	Clarification (juice), flavor (beer), cork stopper treatment Denim finishing, cotton softening De-sizing Scouring Bleach termination Bleaching Excess dye removal
Pulp and paper	→ Lipase Protease Amylase Xylanase Cellulase	Pitch control, contaminant control Biofilm removal Starch-coating, de-inking, drainage improvement Bleach boosting De-inking, drainage improvement, fiber modification
Fats and oils	→ Lipase Phospholipase	Transesterification De-gumming, lyso-lecithin production
Organic synthesis	→ Lipase Acyase Nitrilase	Resolution of chiral alcohols and amides Synthesis of semisynthetic penicillin Synthesis of enantiopure carboxylic acids
Leather	Protease → Lipase	Unhairing, bating De-pickling
Personal care	Amyloglucosidase Glucose oxidase Peroxidase	Antimicrobial (combined with glucose oxidase) Bleaching, antimicrobial Antimicrobial

Figure B.2 Liste des principales enzymes utilisées en industrie.

Les lipases sont utilisées dans différents procédés industriels et sont identifiées par une flèche. Tableau tiré de Kirk *et al.*, 2002.

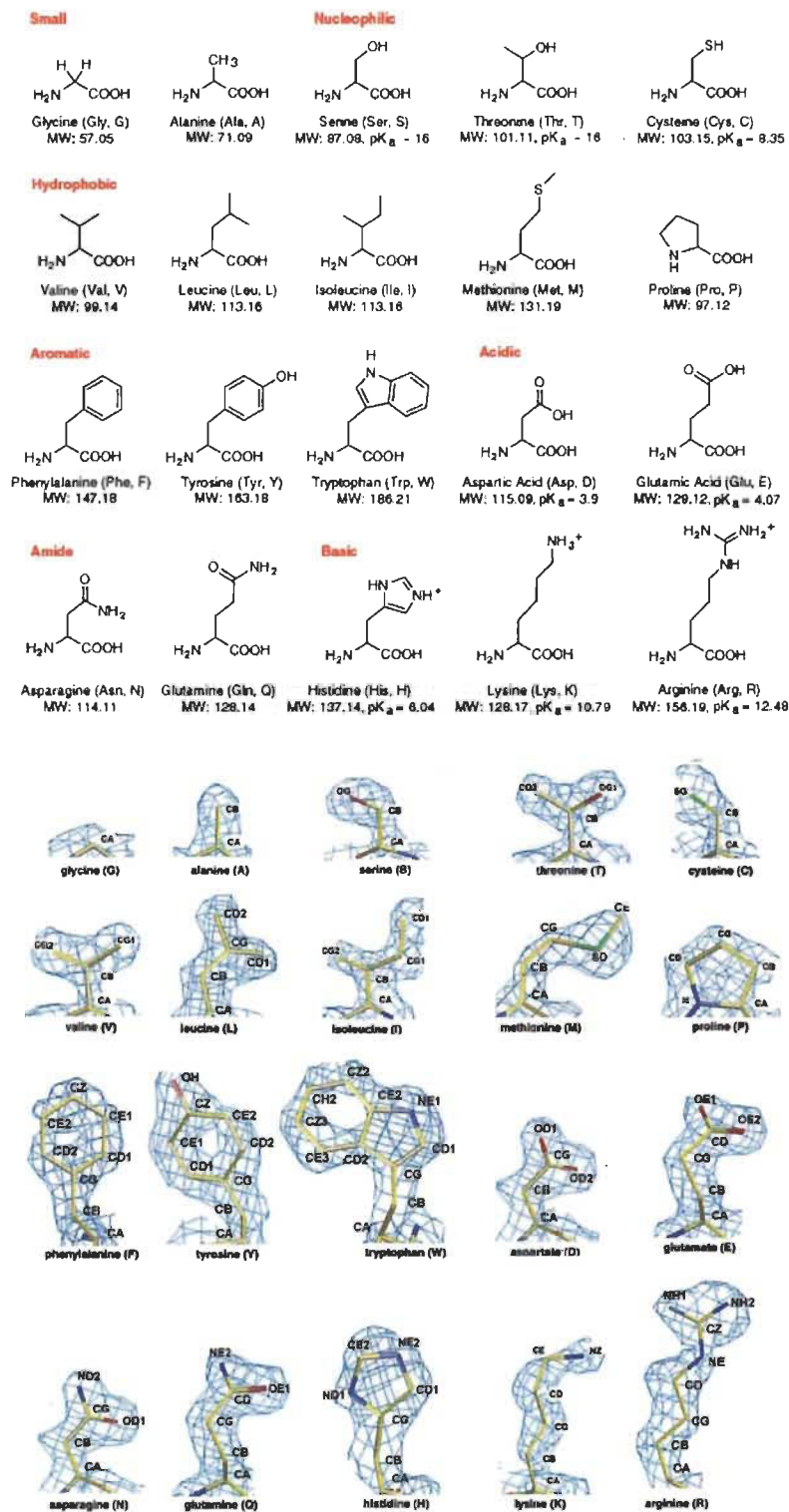


Figure B.3 Structure des acides aminés.

Structure et propriété des acides aminés entrant dans la composition des protéines (haut). Structure et densité électronique des différents résidus d'acides aminés correspondants (bas).

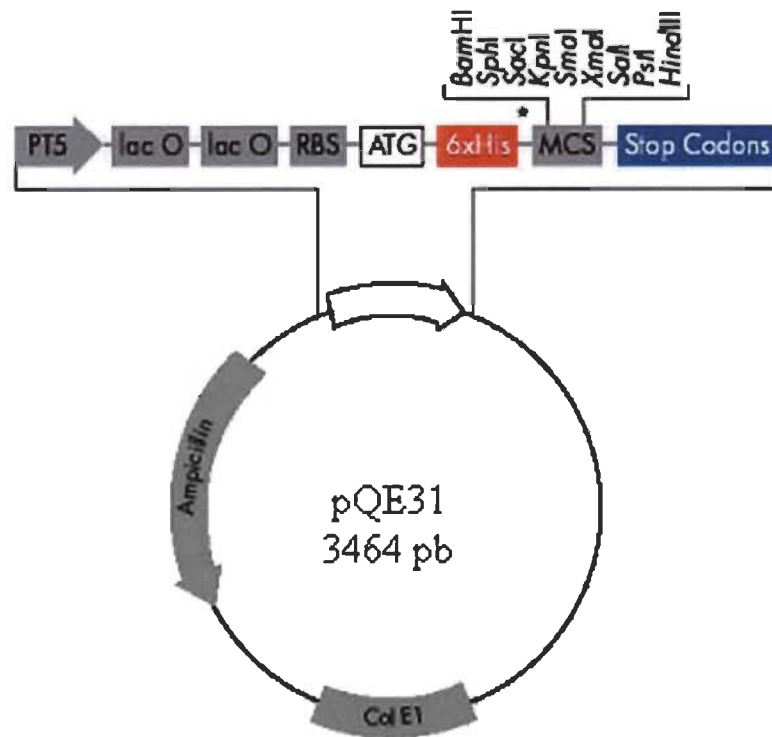


Figure B.4 Vecteur utilisé pour le clonage et l'expression de EstGtA2. L'ORF de EstGtA2 a été inséré entre les sites de restrictions *HindIII* et *KpnI*. La fusion avec une étiquette de purification His₆-tag en N-terminal permet la purification IMAC.

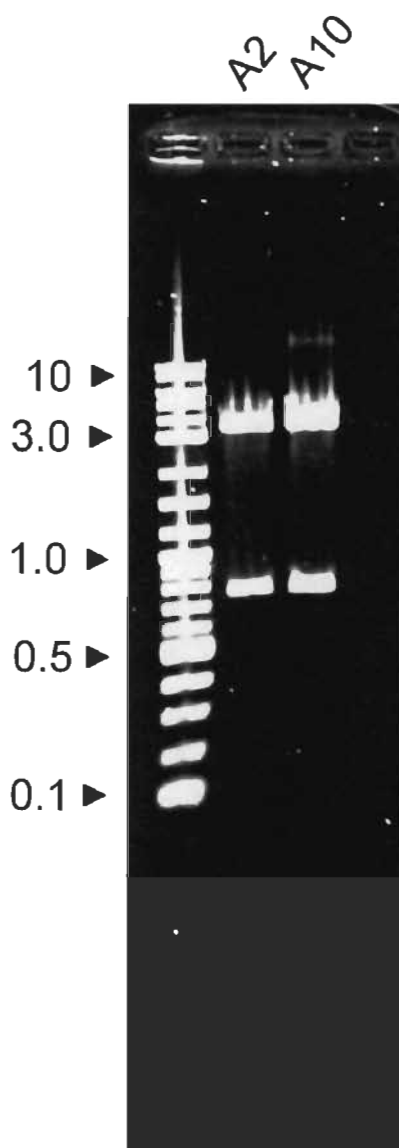


Figure B.5 Double digestion du plasmide recombinant pQE31/estGtA2.

L'ORF de EstGtA2 et EstGtA10 ont été amplifiés par PCR en introduisant les sites de restriction *Hind*III et *Kpn*I à partir de l'ADN génomique de *G. thermodenitrificans* CMB-A2 et CMB-A10 respectivement. Chaque ORF a été inséré dans le plasmide pQE31 (digéré *Hind*III et *Kpn*I), ligué à l'aide de la ligase du phage T4 puis cloné chez *E. coli* M15. Suite à une extraction et à la purification de l'ADN plasmidique recombinant (pQE31/estGtA2 et pQE31/estGtA10), ces derniers sont digérés à l'aide des mêmes enzymes de restriction *Hind*III et *Kpn*I libérant l'insert (ORF de EstGtA2 et EstGtA10). La figure montre les fragments d'ADN suite à une électrophorèse d'ADN en gel d'agarose 0.9 %, 60 min à 120 V. La bande supérieure correspond au plasmide pQE31 linéarisé (3.64 kpb) et la bande inférieure montre l'ORF de EstGtA2 et EstGtA10 (0.747 kpb).

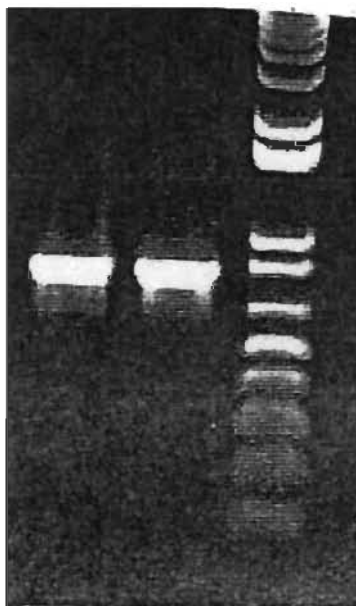


Figure B.6 Amplification de l'ORF de EstGtA2.

L'ORF de EstGtA2 a été amplifié par PCR à partir de la construction pQE31/estGtA2 en insérant les sites de restriction *NdeI* et *EagI* pour le clonage directionnel dans pET28(a+). La figure montre les produits de PCR suite à une électrophorèse d'ADN en gel d'agarose 0.9 %, et à une migration de 60 min à 120 V. Les deux premiers puits contiennent le produit de PCR et le troisième puits contient le marqueur de poids moléculaire (1 Kb Plus DNA ladder).

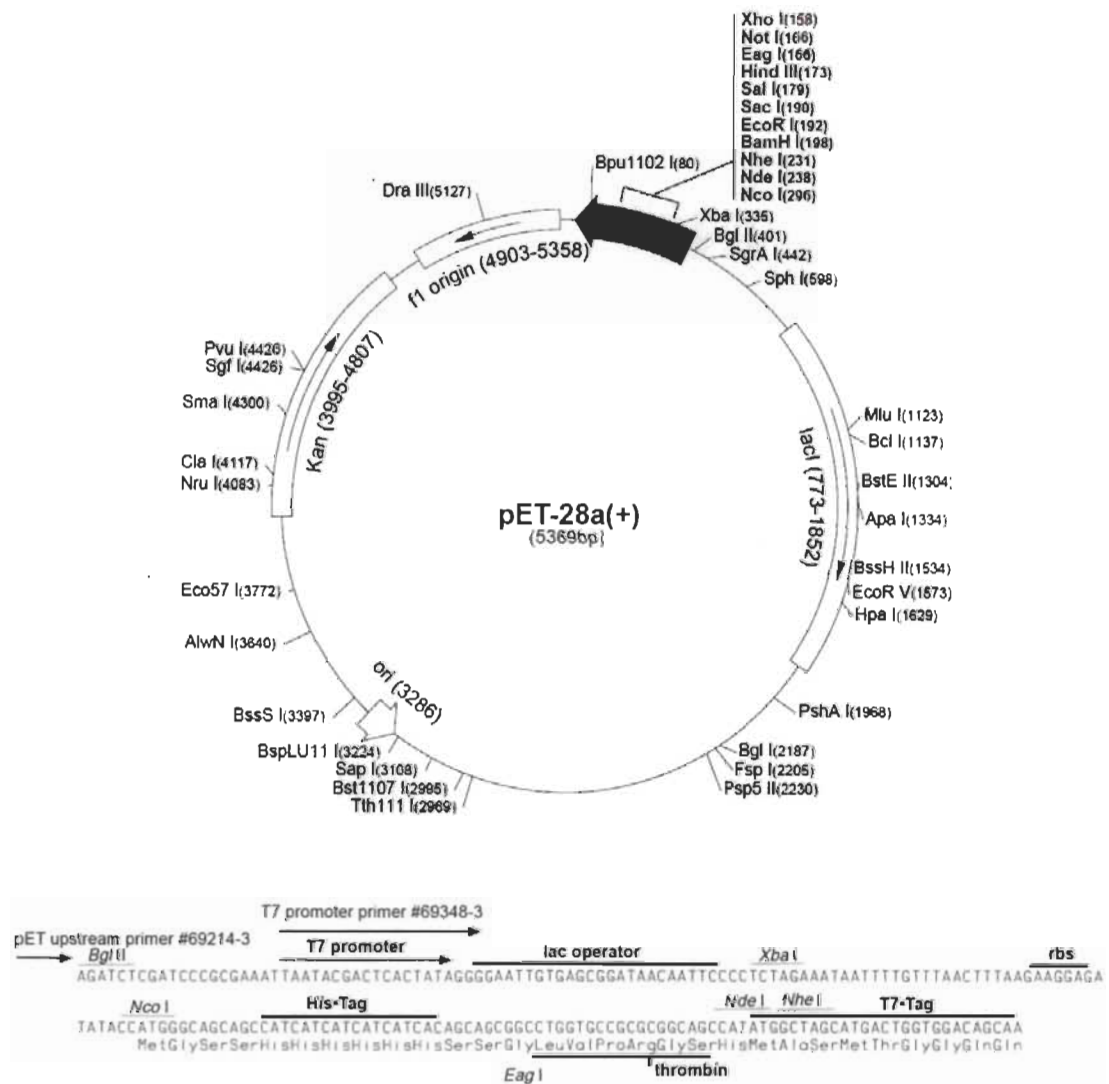


Figure B.7 Vecteur utilisé pour le sous-clonage et l'expression de EstGtA2.
Le produit de PCR de la figure B.6 a été inséré dans le vecteur pET28(a+) entre les sites de restriction *NdeI* et *EagI*. La fusion avec une étiquette de purification His₆-tag en N-terminal permet la purification IMAC. Un site de clivage par la thrombine permet l'élimination du « tag » après la purification.



Figure B.8 Colonies PCR.

Le produit de PCR de la figure B.6 a été inséré dans le vecteur d'expression pET28(a+) et cloné chez *E. coli* XL1-Blue et *E. coli* BL21 Gold. Un total de 5 clones chez *E. coli* XL1-Blue et 8 clones chez *E. coli* BL21 Gold a été sélectionné et la présence de l'insert vérifiée par colonie PCR. La figure montre les produits PCR suite à une électrophorèse d'ADN en gel d'agarose 0.9 %, et à une migration de 60 min à 120 V.

```

1  atgggcagca gccatcatca tcatcatcac agcagcggcc tgggtgccgc cggcagccat
   m g s s h h h h h s s g l v p r g s h

61  atgaaagaac gatatcctgt acttcccggc gcagagccgt tttacgctga aaacggggccg
   m k e r y p v l p g a e p f y a e n g p

121 gtcgggggtgc tgttgctgca cgggttcacg ggcacgccat acagcatgcg tccgcttgct
   v g v l l s h g f t g t p y s m r p l a

181 gaagcgtatg cccaagccgg ttataccgtt tgcttgccgc gottaaaggg gcatgggacg
   e a y a q a g y t v c l p r l k g h g t

241 cactacgaag acatggaacg aacgacgttc cacgattgga ttgcctcggc cgaggaagga
   h y e d m e r t t f h d w i a s v e e g

301 tatgaatggc tgaacaacg atgccaaacc atttttgtca cggggctgtc gatgggtggg
   y e w l k q r c q t i f v t g l s m g g

361 acgctcacgc tttatttggc ggaacagcac ccggagattt gccgcatcgt gccgattaac
   t l t l y l a e q h p e i c g i v p i n

421 gccgctgtcg atattccggc catgcgcgct gggatgacgg gtggtggcga ggtgccgagg
   a a v d i p a l a a g m t g g g e v p r

481 tatcttgact cgatcgggtc agatttgaaa aatccagatg tgaaagaatt atcatagcag
   y l d s i g s d l k n p d v k e l s y e

541 aaaacaccaa ctgcacgct ccttcacgtg gctcaattga tggagcgtgt gaaagaagag
   k t p t a s l l q l a q l m e r v k e e

601 cttggacgca ttacttgccc ggcgctaatt ttcgtttccg atgaagacca tgttggtgccg
   l g r i t c p a l i f v s d e d h v v p

661 cctgggaatg ccgatatcat ttttcaaggc gttcaatcgt cagagaaaga gatcgctccgt
   p g n a d i i f q g v q s s e k e i v r

721 cttcacaaca gctaccatgt agcaacatta gattatgata agcaaacgat tatcgaacgg
   l h n s y h v a t l d y d q q t i i e r

781 tcacttcagt ttttcgccaa acatgcttca
   s l q f f a k h a -

```

Figure B.9 Séquence de EstGtA2 recombinante.

Séquence d'ADN (ORF) encodant EstGtA2 (61-811), en phase avec la séquence encodant l'étiquette de purification (His₆-tag) de pET28(a+), en N-terminal de EstGtA2 (1-60). Le séquençage de l'ORF de EstGtA2 a été effectué à la plateforme d'analyse génomique de l'Université Laval en utilisant l'amorce du promoteur T7 (T7+) et celle du terminateur T7 (T7-).

T (°C)	T (K)	F_a	T_e (K)	$CD_{222\text{ nm}}$	y_F	y_D	fd	K	ΔG
40	313.15	1.92	311.23	-32.0673	-32.43879	-24.93666	0.0495182	0.052098	1827.1842
41	314.15	1.97	312.18	-32.3027	-32.33591	-24.81050	0.0044125	0.0044321	3361.345
42	315.15	1.84	313.31	-32.2541	-32.21353	-24.66043	-0.005372	-0.005343	-
43	316.15	1.93	314.22	-32.1982	-32.11497	-24.53958	-0.010986	-0.010867	-
44	317.15	1.95	315.20	-32.0001	-32.00884	-24.40944	0.0011502	0.0011515	4237.9686
45	318.15	2.05	316.10	-31.9633	-31.91137	-24.28992	-0.006814	-0.006768	-
46	319.15	2.31	316.84	-31.7281	-31.83123	-24.19165	0.0134992	0.0136839	2701.7839
47	320.15	2.67	317.48	-31.7843	-31.76192	-24.10666	-0.002924	-0.002915	-
48	321.15	2.96	318.19	-31.8321	-31.68502	-24.01237	-0.019169	-0.018808	-
49	322.15	3.15	319.00	-31.4302	-31.59730	-23.90480	0.0217223	0.0222047	2413.3655
50	323.15	3.21	319.94	-31.4736	-31.49550	-23.77997	0.0028382	0.0028462	3726.4382
51	324.15	2.96	321.19	-31.4158	-31.36012	-23.61397	-0.007188	-0.007136	-
52	325.15	2.95	322.20	-31.2125	-31.25074	-23.47984	0.0049209	0.0049453	3399.0911
53	326.15	2.98	323.17	-31.2762	-31.14569	-23.35102	-0.016744	-0.016468	-
54	327.15	3.09	324.06	-31.2218	-31.04930	-23.23283	-0.022069	-0.021592	-
55	328.15	3.13	325.02	-30.9027	-30.94533	-23.10534	0.005438	0.0054678	3363.9766
56	329.15	3.13	326.02	-30.7506	-30.83703	-22.97254	0.0109903	0.0111124	2914.9085
57	330.15	3.24	326.91	-30.7441	-30.74065	-22.85435	-0.000438	-0.000438	-
58	331.15	3.38	327.77	-30.4970	-30.64751	-22.74014	0.0190340	0.0194034	2567.5433
59	332.15	3.38	328.77	-30.5520	-30.53921	-22.60734	-0.001613	-0.00161	-
60	333.15	3.39	329.76	-30.4318	-30.43199	-22.47587	2.401E-05	2.401E-05	6969.816
61	334.15	3.40	330.75	-29.8935	-30.32478	-22.34440	0.0540421	0.0571295	1881.1932
62	335.15	3.41	331.74	-29.8677	-30.21756	-22.21293	0.0437068	0.0457044	2033.9006
63	336.15	3.42	332.73	-29.6539	-30.11034	-22.08146	0.0568499	0.0602766	1857.0023
64	337.15	3.43	333.72	-29.3943	-30.00312	-21.94998	0.0756008	0.0817837	1660.1922
65	338.15	3.44	334.71	-29.0770	-29.89591	-21.81851	0.1013826	0.1128206	1451.151
66	339.15	3.45	335.70	-28.4680	-29.78869	-21.68704	0.1630149	0.1947645	1091.247
67	340.15	3.46	336.69	-27.4821	-29.68147	-21.55557	0.2706619	0.3711062	663.16056
68	341.15	3.47	337.68	-26.1218	-29.57426	-21.42410	0.4236060	0.7349245	206.65043
69	342.15	3.48	338.67	-24.2162	-29.46704	-21.29262	0.6423504	1.796033	-394.0589
70	343.15	3.49	339.66	-23.0253	-29.35982	-21.16115	0.7726281	3.3980813	-825.5505
71	344.15	3.50	340.65	-22.0007	-29.25261	-21.02968	0.8819131	7.4683374	-1360.967
72	345.15	3.51	341.64	-21.3233	-29.14539	-20.89821	0.9484562	18.400975	-1977.052
73	346.15	3.52	342.63	-20.8782	-29.03817	-20.76674	0.9865241	73.206567	-2922.893
74	347.15	3.53	343.62	-20.6762	-28.93095	-20.63526	0.9950653	201.64528	-3623.142
75	348.15	3.54	344.61	-20.4923	-28.82374	-20.50379	1.0013813	-724.9771	-
76	349.15	3.55	345.60	-20.3315	-28.71652	-20.37232	1.0048921	-205.4095	-
77	350.15	3.56	346.59	-20.1554	-28.60930	-20.24085	1.0102108	-98.93508	-
78	351.15	3.57	347.58	-20.1932	-28.50209	-20.10938	0.9900124	99.12419	-3174.446
79	352.15	3.58	348.57	-19.9854	-28.39487	-19.97790	0.9991095	1122.0107	-4864.105
80	353.15	3.59	349.56	-19.8331	-28.28765	-19.84643	1.0015795	-634.1073	-

Figure B.10 Mesure des paramètres thermodynamiques de dénaturation.

Exemple de données de dichroïsme circulaire. Mesures et calculs des paramètres thermodynamiques associés à la dénaturation thermique de EstGtA2 à pH 8.

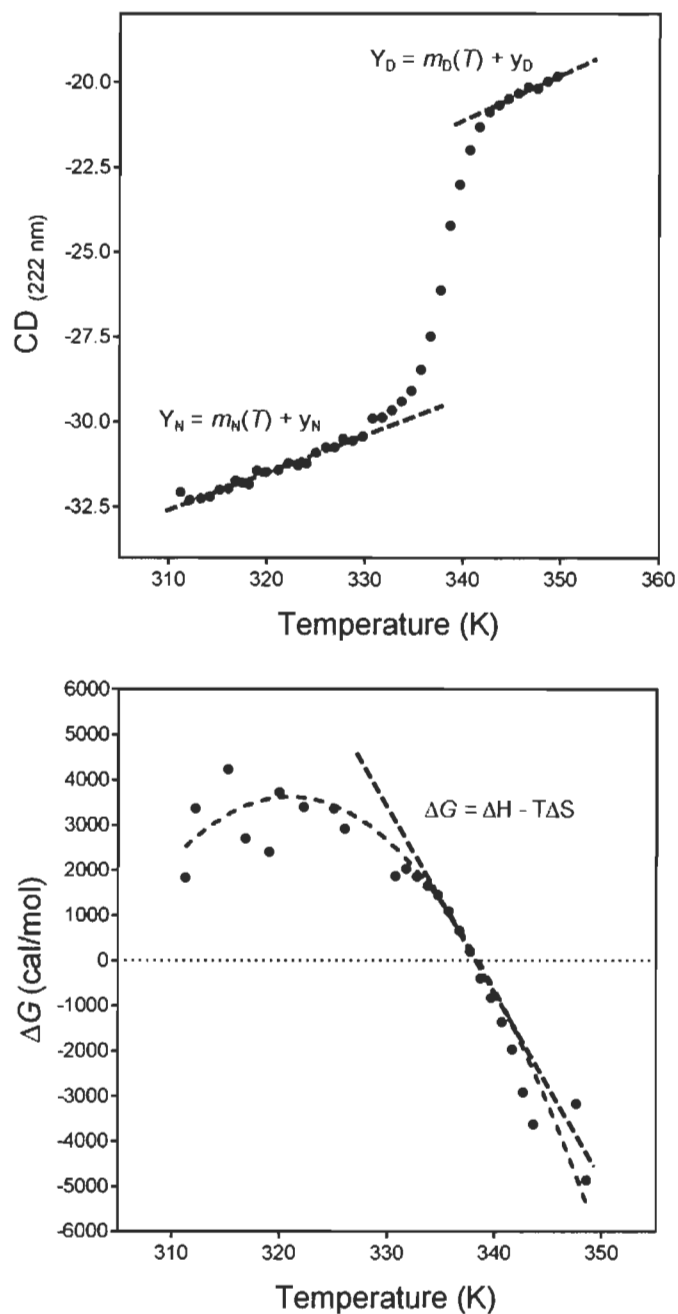


Figure B.11 Dénaturation thermique de EstGtA2 à pH 8 (dichroïsme circulaire).

En haut, exemple d'une courbe de fusion de EstGtA2 à pH 8. Le signal CD à 222 nm est suivi en fonction de la température. Assumant un modèle de dénaturation en deux états (natif et dénaturé), l'équation de la régression linéaire dans la région de pré-transition (native) et de post-transition (dénaturée) est utilisée pour normaliser le signal et calculer la constante de dénaturation, la différence d'énergie libre et la fraction native et dénaturée à chaque température afin d'estimer la température de fusion (T_m) et l'énergie libre de stabilisation à ($\Delta G_{25^\circ\text{C}}$).

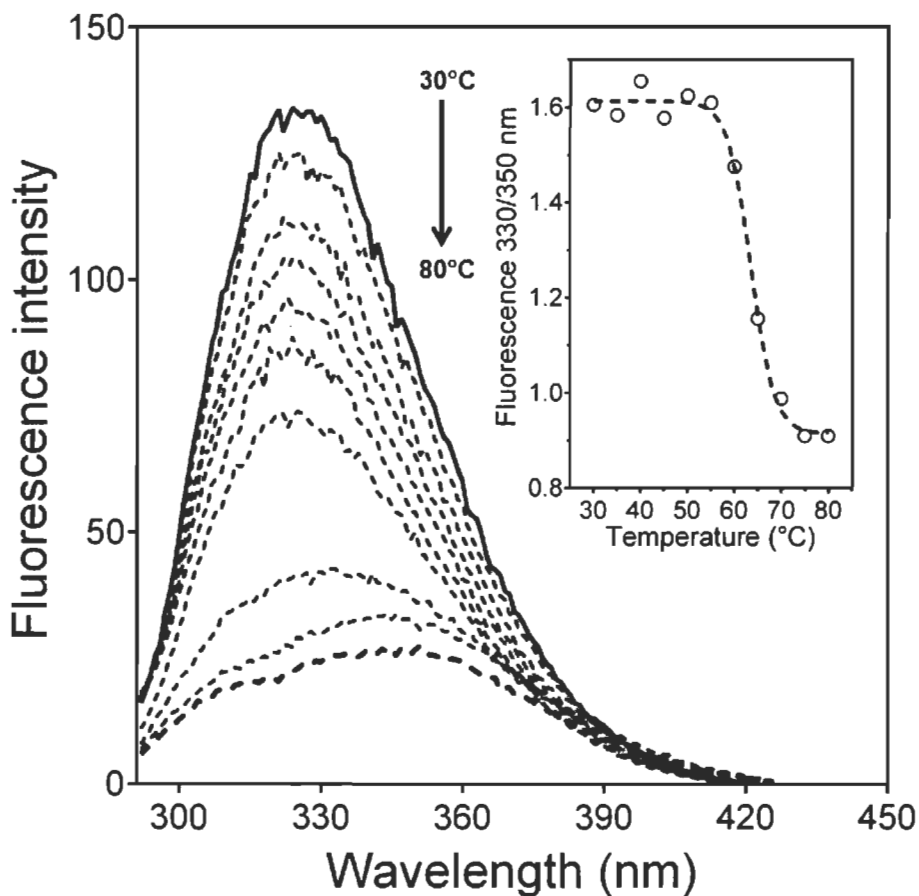


Figure B.12 Dénaturation thermique de EstGtA2 à pH 8 (fluorescence).

Spectres de fluorescence intrinsèque de EstGtA2 mesurés à pH 8 en fonction de la température (30-80 °C). Les spectres d'émission sont mesurés entre 290 et 450 nm suite à une excitation à 287 nm. La longueur d'onde d'émission maximale (λ_{max}) passe de 330 nm (30 °C) à 350 nm (80 °C). L'insertion montre la courbe de fusion, le suivi du ratio entre le signal d'émission de fluorescence à 330 nm (natif) et à 350 nm (dénaturé) en fonction de la température. La température de fusion (T_m) estimée par la courbe de fusion est de 64 °C ce qui concorde avec la valeur de T_m obtenue en CD.

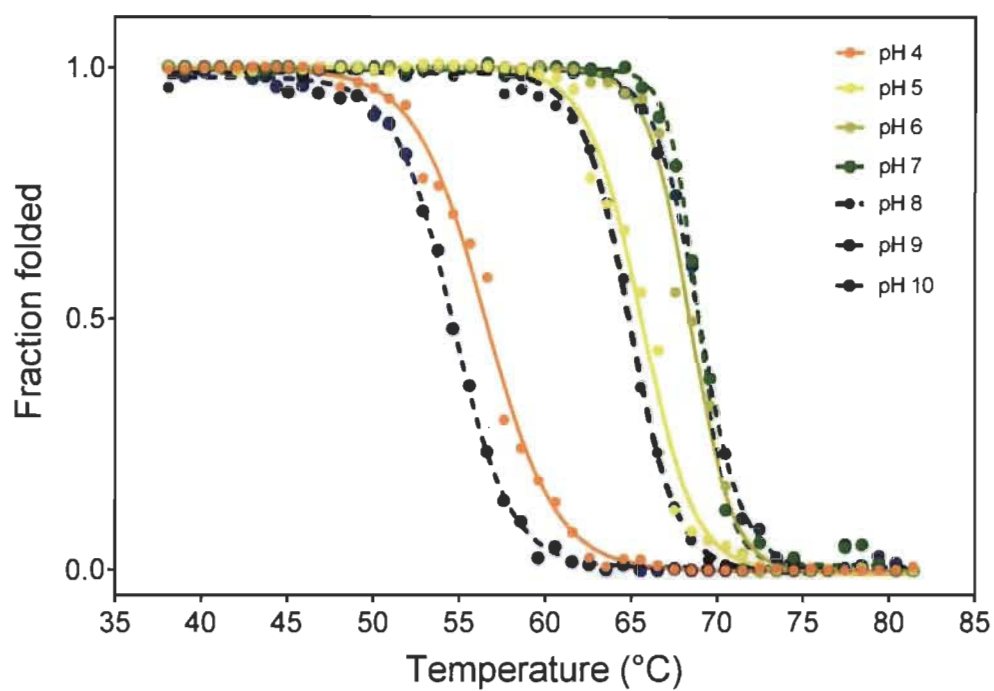


Figure B.13 Dénaturation thermique de EstGtA2 à différents pH.

Les courbes de fusion de EstGtA2 enregistrées à différentes valeurs de pH (pH 4-10). La fraction repliée (dérivée du signal CD à 222 nm) est suivie en fonction de la température et du pH. Les différentes courbes sont montrées selon le code de couleur en fonction du pH identifié.

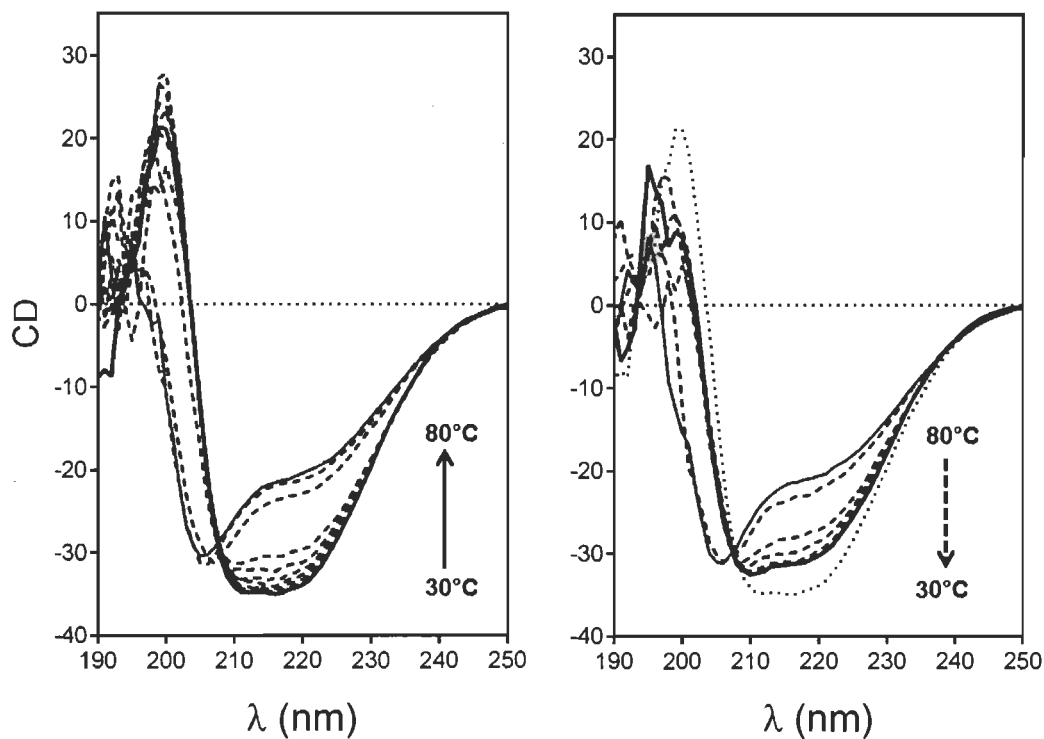


Figure B.14 Réversibilité conformationnelle partielle de EstGtA2 à pH 8.

Spectres CD de EstGtA2 enregistrés à pH 8 en fonction de la température durant le chauffage de 30 à 80 °C (gauche). Spectres enregistrés suite au refroidissement de 80 à 30 °C (droite). Les courbes en gras sont celles enregistrées à 30 °C avant dénaturation (gauche) et après renaturation (droite).

Table II. Dynamic light scattering measurements of EstGtA2: temperature and pH effects.

pH	R_h^a (nm)		MW (kDa)		Polydispersity (%)		mass contribution ^b (%)	
	WT	WT ^c	WT	WT ^c	WT	WT ^c	WT	WT ^c
4	2.5 ± 0.1	n/d ^d	29.3 ± 1.2	n/d	10.9 ± 1.4	n/d	100 ± 0.0	n/d
5	2.4 ± 0.0	n/d	26.7 ± 0.6	n/d	6.1 ± 2.1	n/d	100 ± 0.0	n/d
8	2.5 ± 0.0	3.4 ± 0.2	28.3 ± 1.5	60 ± 0.2	12.2 ± 3.9	11.3 ± 3.8	100 ± 0.1	95.3 ± 0.7
10	2.0 ± 0.1	2.7 ± 0.0	36.3 ± 2.5	34.3 ± 0.6	11.9 ± 0.2	15.2 ± 1.8	99.9 ± 0.0	100 ± 0.1

^a Hydrodynamic radius.

^b Contribution of scattering species to total mass of scattering species in experiment.

^c WT^c = wild-type EstGtA2 after heating to 80 °C followed by cooling to 25 °C.

^d n/d = no detectable scattering species.

Figure B.15 Réversibilité conformationnelle dépendante du pH.

Analyse par diffusion de lumière dynamique (DLS) de EstGtA2 avant et après dénaturation thermique à différents pHs. Les données montrent un état dénaturé soluble en milieu alcalin, à pH 8 et 10, permettant la renaturation et un état dénaturé insoluble en milieu acide, à pH 4 et 5 causant l'agrégation et la déplétion de EstGtA2 ce qui explique la chute du signal CD et la réduction de polydispersité. Données mesurées avec la collaboration de Peter D. Pawelek et Daniel Foshag à l'Université Concordia.

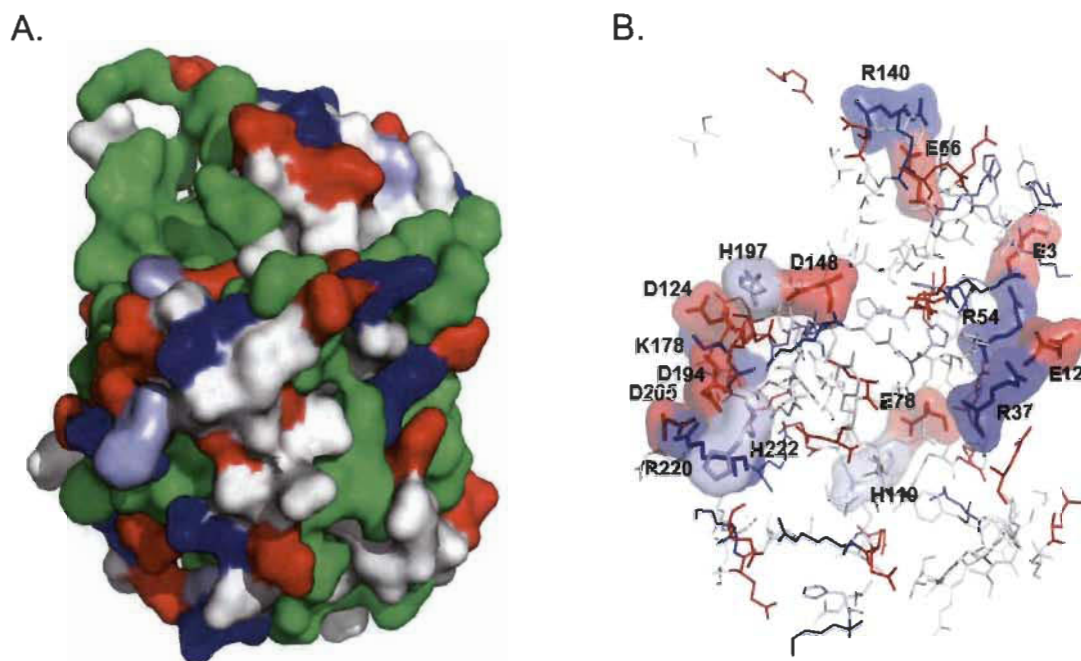
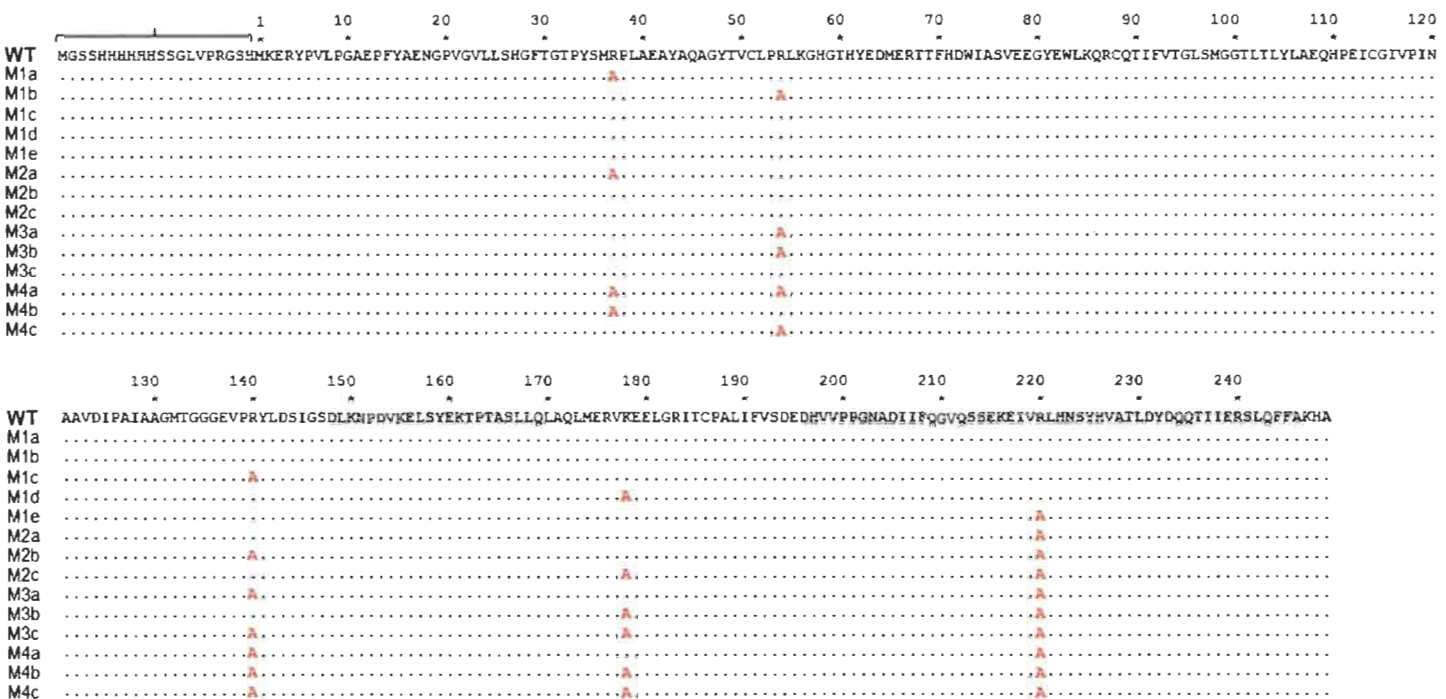


Figure B.16 Ponts salins étudiés dans EstGtA2.

Structure monomérique de EstGtA2. A. La surface des résidus hydrophobes est représentée en vert alors que celle des résidus polaires non chargés en blanc. Les résidus chargés négativement et positivement sont montrés en rouge et en bleu respectivement. B. Résidus polaires. Les chaînes latérales chargées négativement (D et E) et positivement (R et K) sont montrées en rouge et en bleu respectivement. Les résidus histidine (H) sont représentés en bleu pâle. Les résidus étudiés et impliqués dans la formation de ponts salins conservés sont représentés par leur surface et identifiés. Les résidus histidine H110, H197 et H222R ont été substitués par des résidus arginine.

Séquences des mutants (R/K \rightarrow A) combinatoires.

niveau des résidus de ponts salins chargés positivement..



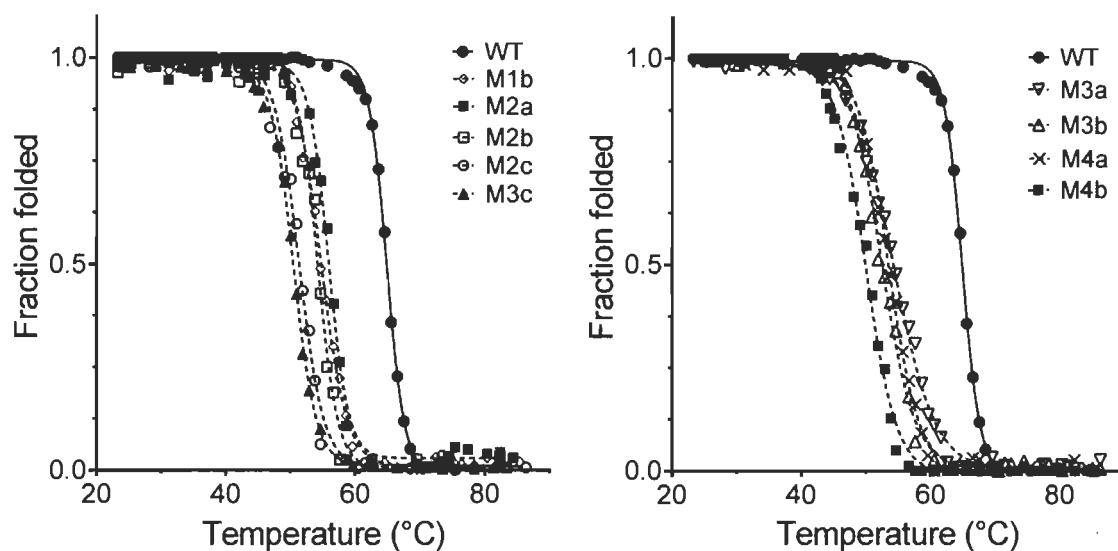


Figure B.18 Impact des substitutions alanine sur la stabilité conformationnelle de EstGtA2.

Courbes de fusion de EstGtA2 (ligne pleine et cercles remplis), ainsi que celles de différents mutants R/K→A. Les courbes montrent la fraction repliée en fonction de la température (25-85 °C) à pH 8.

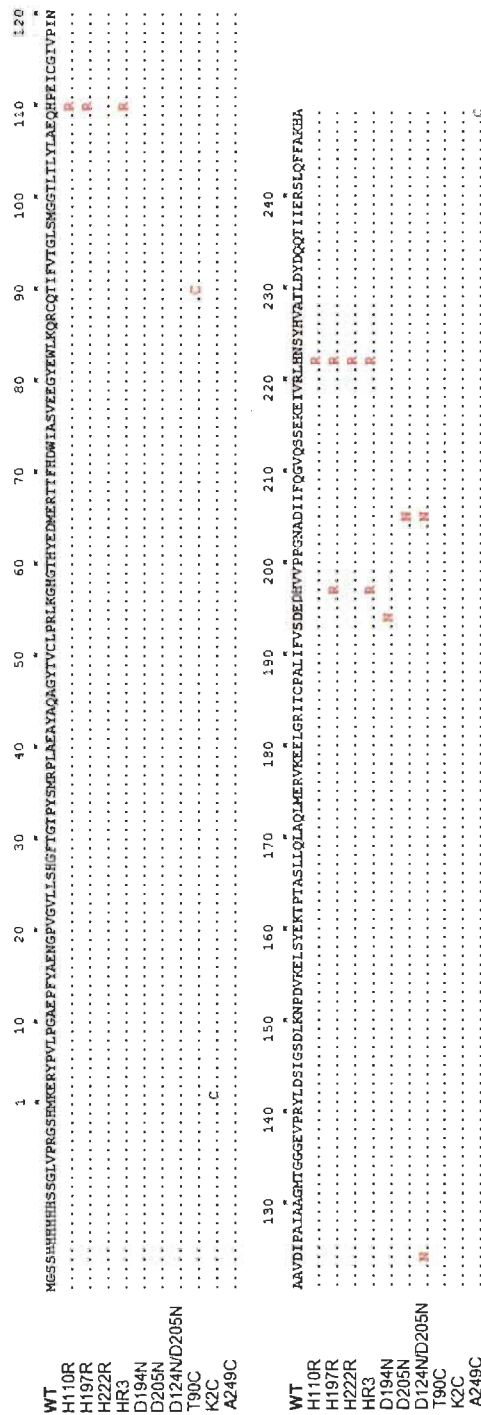


Figure B.19 Séquences des mutants H→R et X→C.

Un alignement entre la séquence du type sauvage de EstGtA2 (WT) et des mutants générés en vue d'augmenter la thermostabilité de EstGtA2 (substitutions H→R) et (T90C). Les substitutions D→N en vue du remplacement isostérique des résidus de ponts salins chargés négativement et les mutants X→C (K2C et A249C) pour la dimérisation artificielle en vue de leur cristallisation.

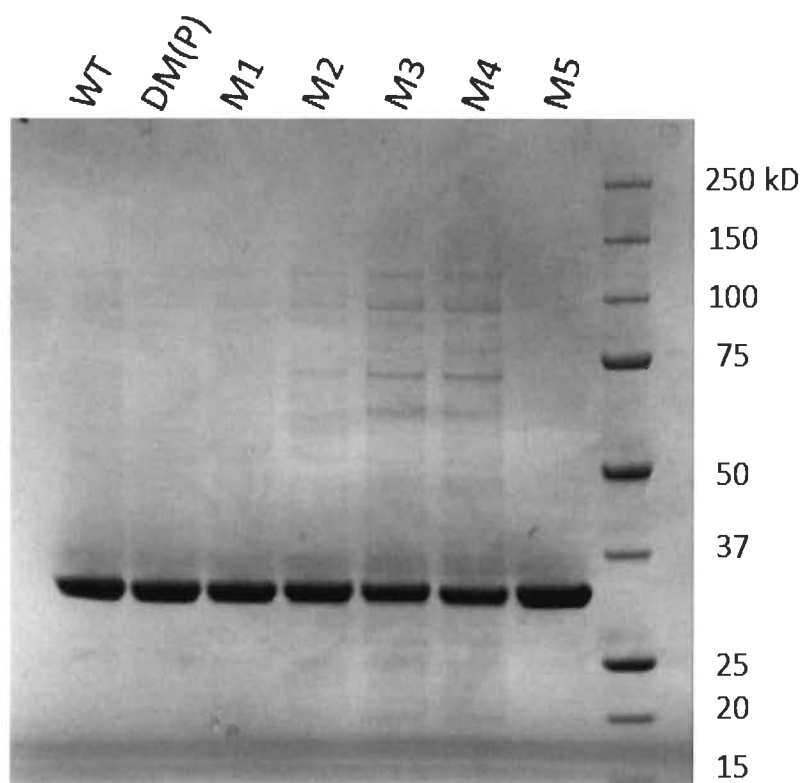


Figure B.20 Gel d'électrophorèse de protéines SDS-PAGE.

Un gel d'électrophorèse du type sauvage de PML et de différents mutants purifiés par chromatographie d'affinité Ni-NTA issus de la première génération d'évolution accélérée. Un total de 10 μ L de chaque protéine (0.5 mg/mL) est déposé sur gel (5 μ g).

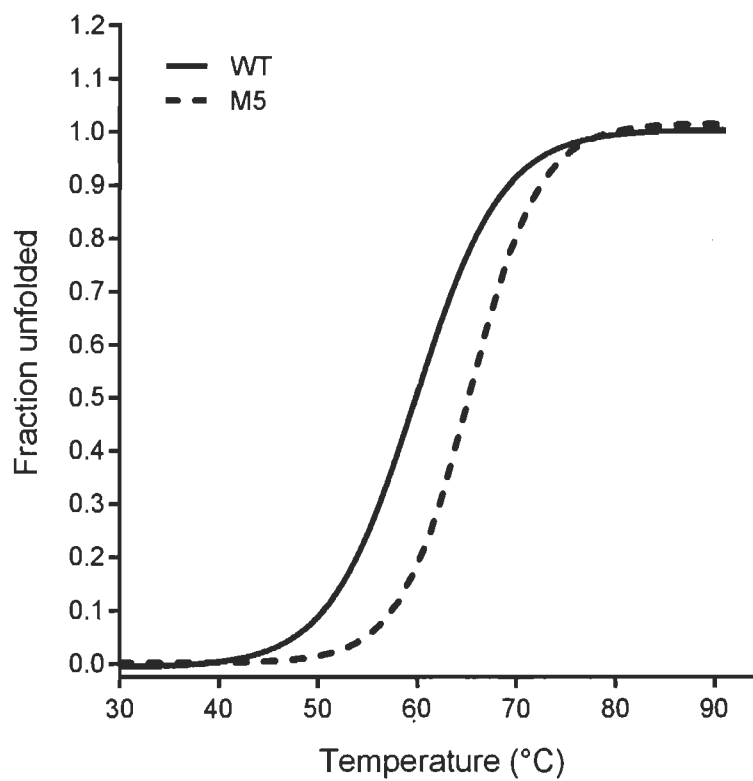


Figure B.21 Courbe de dénaturation thermique de PML et de M5.

Courbe de fusion pour le type sauvage de PML (ligne pleine) et M5 (pointillé). Les courbes de fusion sont dérivées de l'ellipticité molaire à 222 nm et converties en fraction pliée.

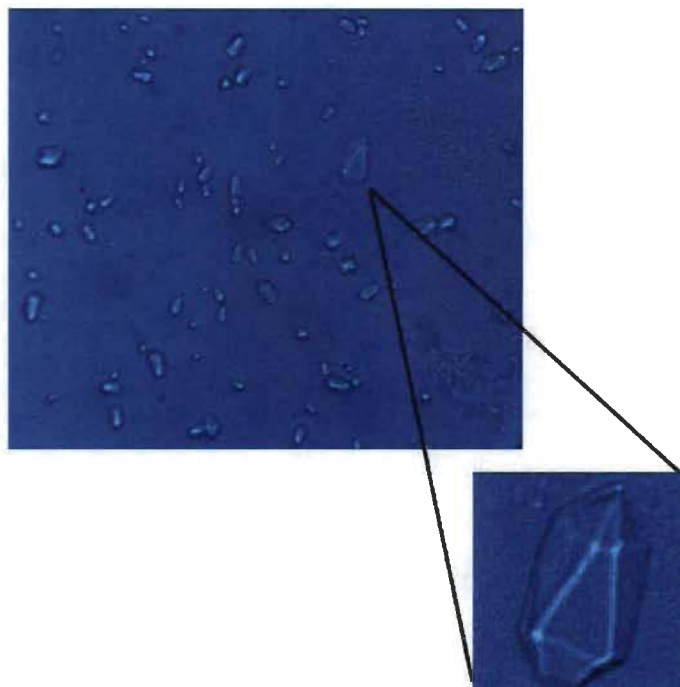


Figure B.22 Cristallisation de M5.

Cristaux de M5 formée après 24 h à 25 °C en goutte suspendue au-dessus d'une solution de cristallisation (liqueur mère). La solution de cristallisation (liqueur mère) contenant 68 % MPD, 1.2 % β -OG, HEPES pH 6.5.

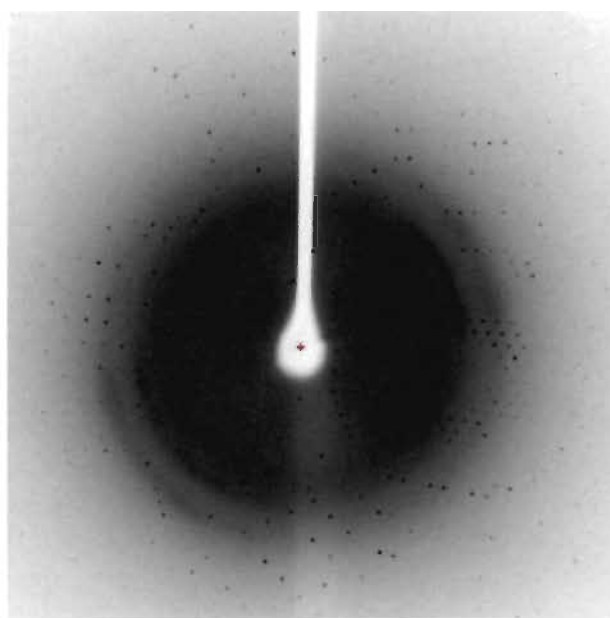


Figure B.23 Patron de diffraction du cristal orthorhombique de M5.

Groupe d'espace : $P3_2$, $a = b = 65 \times 14$, $c = 63$, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$.

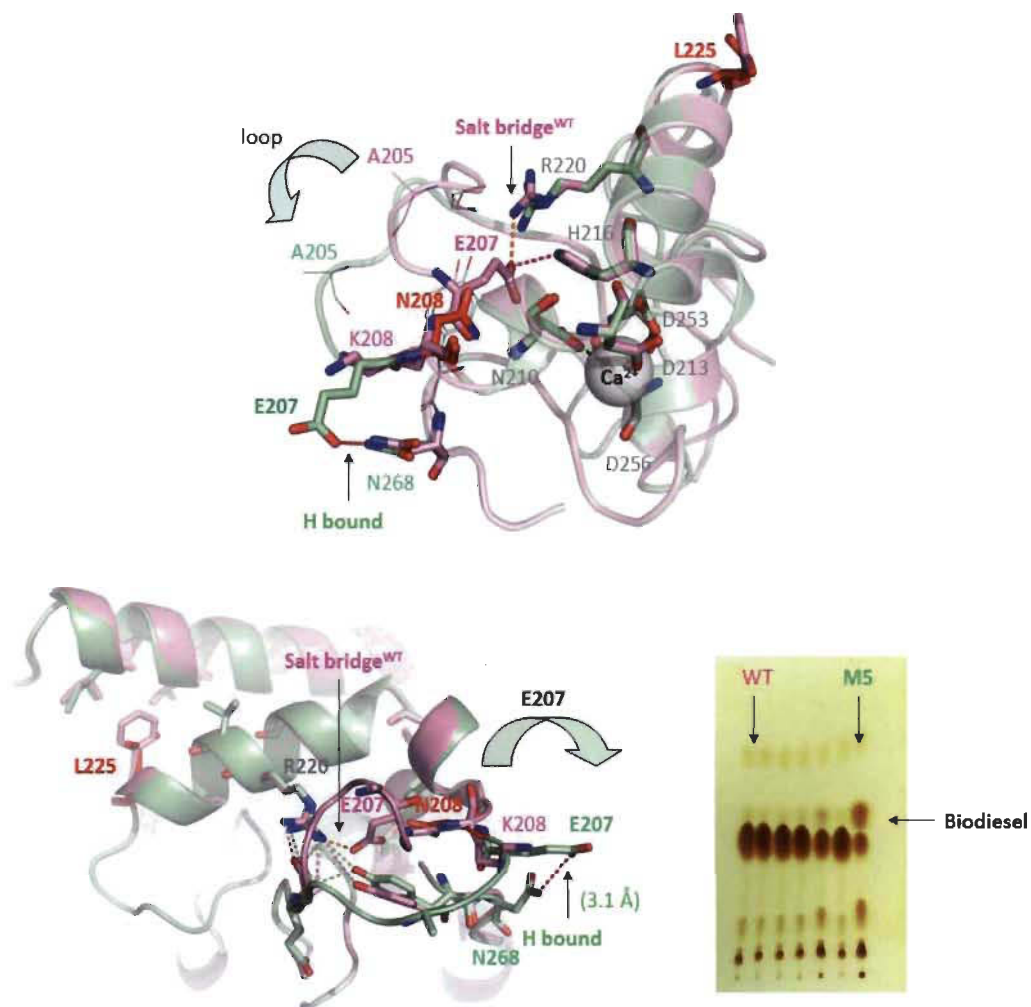


Figure B.24 Changements structuraux dans le mutant M5.

La figure montre une portion de l'alignement de la structure cristalline du type sauvage de PML avec celle de M5. La réorganisation d'une boucle dans l'environnement du site de liaison du calcium, ainsi que la formation de ponts hydrogènes distants semblent être responsables de la thermostabilisation de M5 (un intermédiaire d'évolution accélérée de Dieselzyme 2).

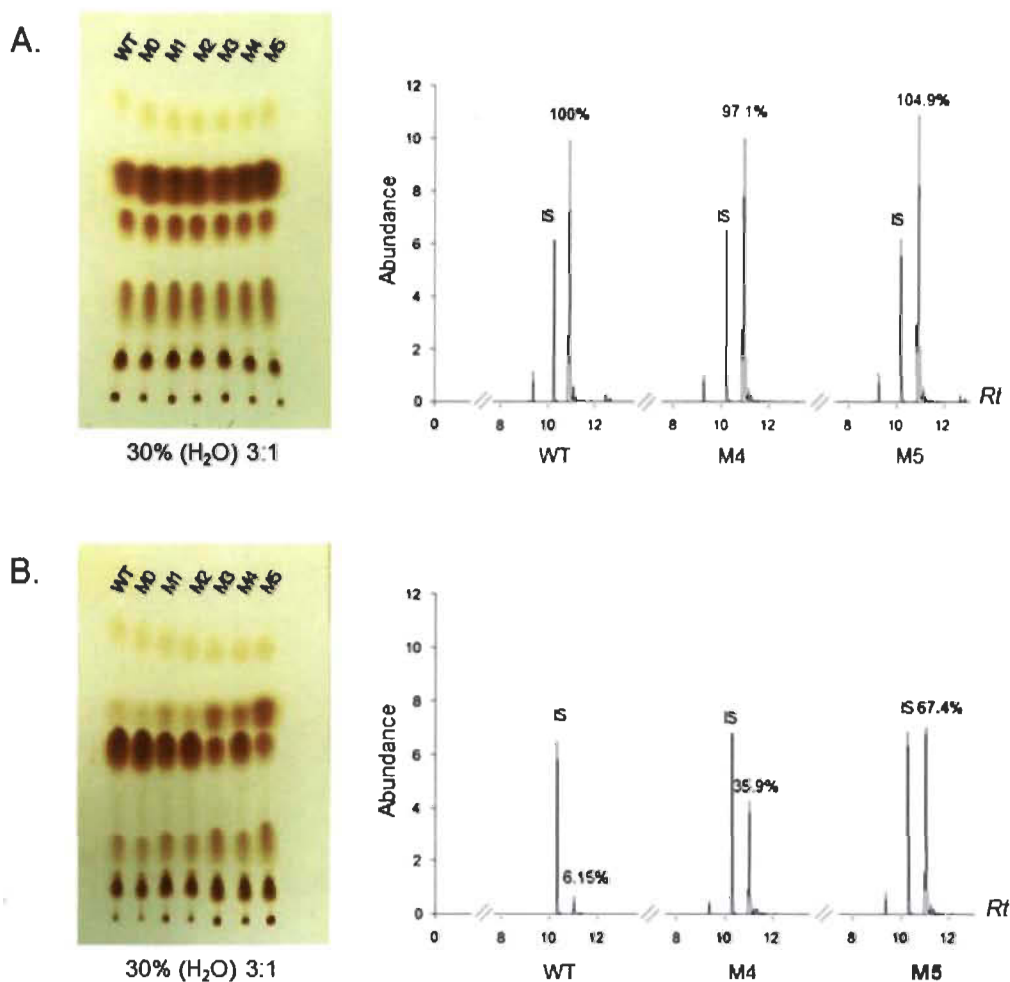


Figure B.25 Activité de transestérification de PML et des différents mutants (première ronde d'évolution accélérée) mesurée par TLC.

Les mutants M0-M5 sont différentes combinaisons des mutations criblées durant la première ronde d'évolution accélérée. La transesterification et la synthèse de biodiesel est détectée suite à une réaction de 20 h à 25 °C. A. Détection par chromatographie sur couche mince (TLC) et par chromatographie en phase gazeuse (GC-FID). La synthèse de biodiesel est effectuée sur une période de 16 h à 25 °C, sans incubation (A) et suite à une incubation de 2 h à 50 °C (B). Chaque réaction contient 30 % d'eau et un ratio molaire méthanol:huile (3 :1), équivalent stoechiométrique. Le pourcentage de synthèse des différents mutants relatif au type sauvage est indiqué au-dessus des pics. Le standard interne est aussi indiqué pour chaque réaction (IS).

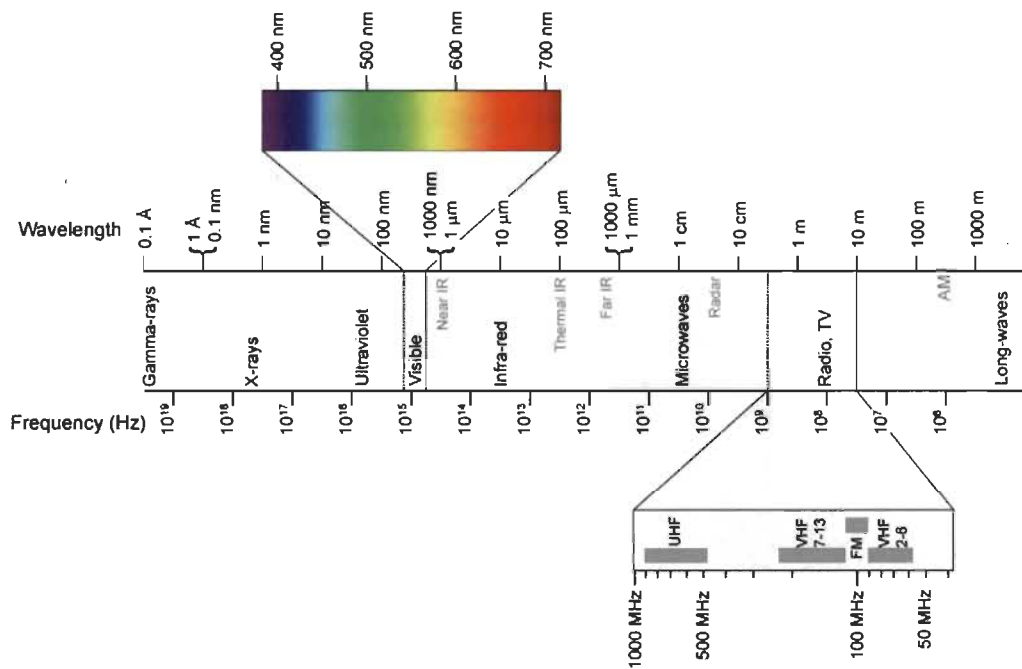


Figure B.26 Spectre électromagnétique de la lumière.

Échelle représentant les différentes radiations électromagnétiques de la lumière. Le spectre du visible correspond à une région étroite du spectre électromagnétique.

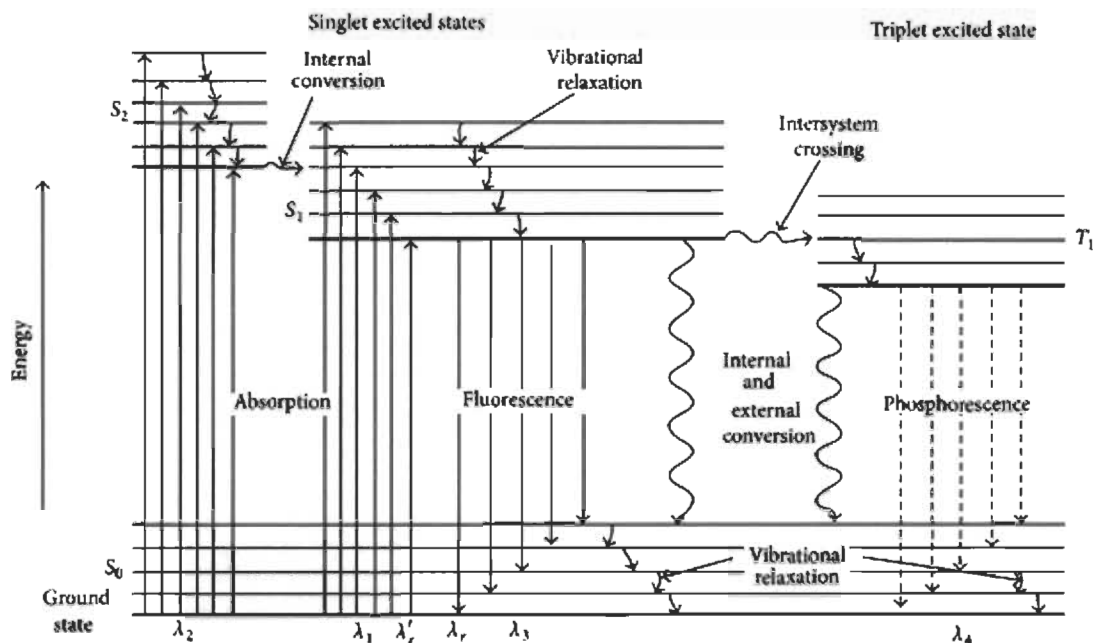


Figure B.27 Diagramme de Jablonski.

Diagramme décrivant les différents niveaux électroniques des molécules organiques et les transitions possibles entre les différents états.